London School of Economics and Political Science

# Evidence, process or context? Examining the factors that drive coverage decisions of pharmaceuticals by Health Technology Assessment bodies in Europe

**Karin H. Cerri**

A thesis submitted to the Department of Social Policy at the London School of Economics for the degree of Doctor of Philosophy, London, December 2011

## Declaration of Authorship

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).



I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

# Abstract

In Europe, Health Technology Assessment (HTA) bodies produce coverage decisions that guide public funding of pharmaceuticals. This thesis examines and weights those factors that drive HTA coverage decisions, focusing on the National Institute for Health and Clinical Excellence (NICE) in England and Wales, the Scottish Medicines Consortium (SMC), the Dutch College voor Zorgverzekeringen (CVZ), and the French Haute Autorité de Sante (HAS). To address the research question, a dataset of approximately 1000 HTA coverage decisions by NICE, SMC, CVZ and HAS from the period 2004-2009 was created, containing more than 30 clinical, economic, process and socio-economic factors extracted from published HTA reports. A three-category outcome variable was used, defined as the decision to 'recommend', 'restrict' or 'not recommend' a technology. Multivariate analyses were conducted to assess the relative contribution of the explanatory variables on coverage decisions both within and between HTA bodies.

Results demonstrate that different combinations of clinical/economic evidence, process and socio-economic factors drive HTA coverage decisions by NICE, SMC, CVZ and HAS. In addition, the same factor may behave differently according to the nature of the coverage decision. The analysis further suggests there is a significant difference between HTA bodies in the probability of reaching a 'restrict' or 'not recommend' decision outcome relative to a 'recommend' outcome, adjusted for evidence, process and context factors. This thesis contributes to the understanding of factors driving HTA coverage decisions by examining multiple European HTA bodies, enhancing the comprehensiveness of the factors examined through descriptive and multivariate analyses and by identifying and weighting the key drivers of the coverage decisions made by the four HTA bodies between 2004 and 2009. This research further provides relevant insights to variation among HTA bodies in the determination of patient access to pharmaceuticals, and implications for collaboration between European HTA bodies.

## Acknowledgements

I would like to thank Martin Knapp and Jose-Luis Fernandez for their guidance and advice in the preparation of this thesis. For their support and patience, I would like to thank and recognise my family, especially Bernadette Cerri and Thomas Morel. For editorial assistance, I would like to thank Anna Maresso. I would like to acknowledge my managers and Bristol-Myers Squibb for giving me the opportunity to pursue this PhD, particularly Klaus Pugner and Peter Hertzman.

# Contents

# List of Acronyms

| | |
|---|---|
| ABPI | Association of the British Pharmaceutical Industry |
| ACD | Appraisal Consultation Document |
| Afssaps | Agence française de sécurité sanitaire des produits de santé |
| ACE inhibitors | angiotensin-converting enzyme inhibitors |
| AIFA | Agenzia Italiana del Farmaco |
| ASMR | Amelioriation du Service Medicale Rendu |
| BBC | British Broadcasting Corporation |
| BNF | British National Formulary |
| BSR | British Society of Rheumatology |
| CDR | Canadian Drug Review |
| CEPS | Comité Economique des Produits de Santé |
| CEA | Cost Effectiveness Analysis |
| CFH | Commissie Farmaceutische Hulp |
| CVZ | College voor Zorgverzekeringen |
| CI | Confidence Interval |
| CML | Chronic Myeloid Leukemia |
| CNS | Central Nervous System |
| CUA | Cost Utility Analysis |
| CVZ | College voor zorgverzekeringen |
| DALE | Disability Adjusted Life Expectancy |
| DEMESP | Direction de l'Evaluation Medicale, Economique et de Sante Publicque |
| EEC | European Economic Community |
| EMA | European Medicines Agency |
| EU | European Union |
| EUnetHA | European Network of HTA |
| EFPIA | European Federation of Pharmaceutical Industries and Associations |
| FAD | Final Appraisal Determination |
| FDA | Food and Drug Administration |
| DH | Department of Health |
| NHS | National Health Service |
| HAS | Haute Autorité de Sante |
| HR-QOL | Health Related – Quality of Life |

| | |
|---|---|
| HTA | Health Technology Assessment |
| ICER | Incremental Cost Effectiveness ratio |
| IHTA | International Health Technology Assessment |
| INAHTA | International Network of Agencies for Health Technology Assessment |
| INAMI | Institut national d'assurance maladie-invalidité, Belgium |
| IRP | Independent Review Panel, SMC |
| ISD | Information Services Division |
| GDP | Gross Domestic Product |
| GVS | Geneesmiddelenvergoedingssysteem |
| GBP | Pound sterling |
| OECD | Organisation for Economic Co-operation and Development |
| PET | Positron emission tomography |
| PBAC | Pharmaceutical Benefits Advisory Committee |
| PCT | Primary Care Trust |
| PhRMA | Pharmaceutical Research and Manufacturers of America |
| MTA | Multiple Technology Appraisal |
| NICE | National Institute for Health and Clinical Excellence |
| NDC | New Drugs Committee |
| NOS | Nederlandse Omroep Stichting |
| NPAF | New Product Assessment Form |
| NZa | Nederlandse Zorgautoriteit |
| QALY | Quality Adjusted Life Year |
| RCF | Rarer Cancers Forum |
| RCT | Randomised Controlled Trial |
| SMC | Scottish Medicines Consortium |
| STA | Single Technology Appraisal |
| Anti-TNF | Anti-tumour Necrosis Factor |
| TAR | Technology Appraisal Report |
| TA | Technology Appraisal |
| USA | United States of America |
| UK | United Kingdom |
| VBP | Value-Based Pricing |
| Zvw | Zorgverzekeringswet |

## Conflict of Interest Statement

I, Karin Cerri, declare that during the time of writing this thesis, I was an employee of Bristol-Myers Squibb (B-MS) within the European Health Economics and Outcomes Research (HEOR) department.

As Associate Director within the HEOR department, my work involved generating and packaging clinical and economic data to support access to B-MS medications within a European treatment setting. In particular, a key task in which I participated was the creation of HTA submissions detailing the clinical and economic characteristics of B-MS medications to reimbursement/HTA bodies across a number of European Member States.

In the context of the thesis, my role at B-MS did not have undue influence on the analysis and research performed, and the opinions and ideas expressed in this work are entirely my own.

# 1 Introduction

> "Our job is to make sure taxpayers' money is only spent on healthcare that works and is good value" (NICE statement on BBC Breakfast News, 13th May 2008)

> "It's a system of blocking. They're [NICE] not looking at patients and saying "how can we fund it?", they are saying, "how can we not fund it?"…[Cancer] medicines are licensed and working yet we in Britain aren't allowed to access them. While we wait for Nice to decide, patients are dying." (Kate Spall1 on BBC Breakfast News 13th May 2008)

The quotations above juxtapose different perspectives on the public funding of pharmaceutical technologies within the healthcare system, raising interesting questions about how funding decisions are made, and the impact this may have on patients, as well as on providers, manufacturers and health policy makers. In the heavily regulated European pharmaceutical market (Maynard and Bloor 2003; Mossialos et al. 2004), one of the areas under considerable regulation is the public funding of pharmaceutical technologies. Health Technology Assessment (HTA) is a process that exists in several European Union (EU) Member States to advise healthcare systems on the appropriate use of a new technology and whether it should be recommended for public funding. Examples of HTA agencies include the National Institute for Health and Clinical Excellence (NICE) in England and Wales, the Scottish Medicines Consortium (SMC), the Commissie voor Zorgverze (CVZ) in the Netherlands and the Haute Autorité de Sante (HAS) in France.

This chapter aims to set the scene for this thesis by presenting the context and rationale for the research question, and by setting out how the thesis is structured. Firstly, the extent and nature of the regulation of the pharmaceutical market in EU Member States is analysed and the concept of HTA is introduced. Coverage decisions and their implications for patients, providers and manufacturers are highlighted in the third

---

[1] Kate Spall is a member of the public who has been involved in helping more than 50 patients across England and Wales to receive funding from local Primary Care Trusts ( PCTs) for new cancer drugs that had not yet been reviewed by NICE.

section of this chapter, and finally the research question is outlined, along with an outline of the chapters that lie ahead.

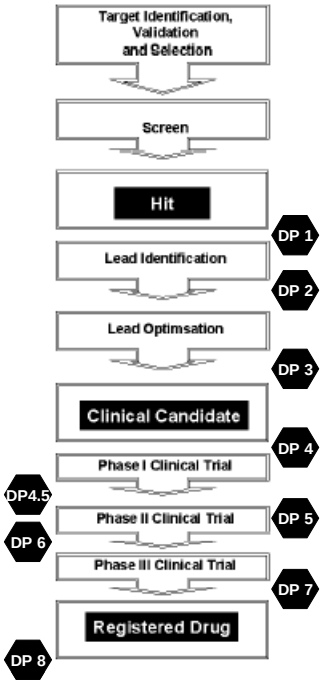
## 1.1 Overview: Regulation of the Pharmaceutical Market in EU Member States

The pharmaceutical market in the EU is heavily regulated for several reasons. Firstly, the presence of regulatory mechanisms to control the level of drug utilisation reflects the oligopoly structure of the pharmaceutical industry and its tendency for monopoly-like power, as well as the characteristics of health system users and its various providers (physicians etc) (McGuire et al. 2004). Secondly, regulation of pharmaceuticals is necessary due to the fact that patients lack complete information about treatment options, and that both patients and providers (due to their 'third party' status) are seldom in a position to bear the full cost of utilisation of pharmaceuticals (Mossialos et al. 2004; Vogel 2004). In addition, other factors have led to the increased involvement of national governments and supra-national governing bodies (i.e. the European Commission) in the regulation of access to healthcare. These include the existence of complex stakeholder relationships, i.e. the provider/payer/patient 'triangle' that characterizes healthcare systems (Reinhardt 1990), the 'imperfect' nature of the healthcare market (Abel-Smith and Grandjeat 1978; Jacobzone 2000; Barr 2001; Dukes 2003; McGuire et al. 2004), and the shift towards a greater use of market forces in the distribution of healthcare (Rice et al. 2000).

Such regulation impacts on the entire 'life-cycle' of pharmaceuticals along their path from discovery in the laboratory to patient use. There are a series of key requirements which must be satisfied, both before and after marketing authorisation has been granted. Such requirements can be categorised into pre- and post-marketing authorisation requirements and are designed to impact on patients, providers, payers and the industry (either individually, or in combination). Pre-marketing authorisation requirements mark key milestones in the discovery and development process of a compound – a process in which, for example, one out of every 10,000 potential medicines investigated by research-based pharmaceutical companies makes it through the research and development pipeline and is approved for patient use by the United States Food and Drug Administration (PhRMA 2007), amounting to an average of 15 years of research

and costing US$800 million (or around €552-€690 million)[2] (European Commission 2009). Pre-marketing authorisation requirements cover various stages of the discovery and development phase, and consist of nine key decision points. These decision points are accurately illustrated by Garret et al. (2003), who provide an analysis of the process with which an anti-tumour target is identified, validated and selected, developed into a clinical candidate, and finally, developed into a registered medication for cancer therapy (Figure 1.1).

**Figure 1.1 Drug discovery and development process**




*Source*: Adapted from Garret et al. 2003.

*Note*: DP - decision point

In contrast to the relative homogeneity of the pre-marketing authorisation requirements across products and geographies, Member States differ in the types of requirements that they have put in place to manage patient access to medications once marketing authorisation is granted. After marketing authorisation, a series of nationally-led payer, provider and patient-related requirements determine patient use and access to pharmaceuticals. In response to specific welfare and healthcare systems, Member

[2] While there is some debate as to the magnitude of the cost of drug development (Permanand 2006; EU Commission 2009), the literature is in agreement on the fact that the 'entry' costs into the market for the pharmaceutical companies are high.

States have adopted distinctive combinations of payer-led requirements (Pricing and Reimbursement (P&R), budget allocation, Heath Technology Assessment), tools aimed at providers (clinical guidelines, gate-keeping mechanisms) and tools aimed at patients (co-payments for medications, doctor's visits to obtain prescription). The next section focuses on payer-led funding requirements in EU Member States to regulate the funding of pharmaceuticals, with a particular focus on Health Technology Assessment (HTA).

## 1.2 Funding requirements in Europe and the role of HTA

> "The [Dutch] government policy aimed at reducing drug expenditure appears to bear fruit. Total drug expenditure increased this year by 2.6 percent to 5.2 million [euros] per year. In 2007 the increase was three times as high." (NOS 2008, p1)

The EU, with its 27 Member States, spends €138 billion annually on pharmaceuticals (EU Commission 2009). Regulations to control and optimise public spending on pharmaceuticals exist in all Member States, initiated and implemented by those entities within the healthcare system that manage pharmaceutical budgets and expenditure, often referred to as payers. Prior to examining the use of such mechanisms and the variation between Member States in their use, the concepts of 'payers' and 'funding requirements' are outlined.

Payers represent the actors within the healthcare system that are responsible for budget allocation and management, either directly, by being accountable for a budget, or indirectly by providing guidance on the allocation of budgets to pharmaceuticals[3]. The term 'payers' represents a heterogeneous pool of actors within the healthcare system. Payers can exist at each level of the healthcare system, whether national, regional or local. National-level payers are those entities, such as Ministries of Health, that hold a budget for pharmaceutical expenditure at the national level. Reimbursement committees (e.g. the INAMI in Belgium) are also examples of national-level payers – while not necessarily having direct accountability for the pharmaceutical budget, they issue guidance on which pharmaceuticals should be covered by the healthcare system. A further example, in Italy, is the authority for evaluation of pharmaceutical products at

[3] While noting that requirements for patient copayment for medication exist in several Member States, for the purposes of this research, patients are not considered as payers.

the national level (AIFA), which grants price approval, determines reimbursement criteria, and is responsible for liaising with the Ministry of Health to publish the price and reimbursement criteria in the national *'Gazzetta'*. Regional level payers are those that hold budgets or have responsibilities for pharmaceuticals at a sub-national level. For example, in Italy, which is comprised of 20 regions, each region has independent responsibility for its own pharmaceutical budget and expenditure. Finally, there are local level payers such as hospital formulary committees, hospital pharmacists, or hospital budget holders that hold responsibility for funding and providing access to pharmaceuticals within their local jurisdiction. To continue the Italian example, the hospitals within each region (assuming a technology is a compound destined for use in a hospital setting), has its own individual procedures for budget allocation and formulary decisions. The combined pool of payers within the healthcare system is collectively responsible for the purchasing of pharmaceutical technologies, aided by the use of funding regulations.

Regulations on public funding of pharmaceuticals, depending on the perspective adopted, can be labelled differently. From a manufacturer perspective, regulations linked to the funding of compounds have been defined as '4$^{th}$ hurdles' in the literature (Maynard and Bloor 2003; Taylor et al. 2004). From the perspective of the health care system, these regulations are not perceived as 'hurdles' but as 'tools' to aid the achievement of system objectives, including assurance of the quality of services provided, and the efficiency of drug utilisation (Saltman et al. 1998; Mrazek 2002; Maynard and Bloor 2003; Mossialos et al. 2004). Whether labelled from a manufacturer or health care system perspective, the attributes of these regulations remain the same.

In an effort to adopt a 'neutral' label for regulations that aim to control and guide public funding of pharmaceuticals, for the purposes of this thesis, these regulations can be considered to be 'funding requirements'. These represent conditions/regulations that must be met for coverage to be provided, and include activities such as the requirement for health technology assessment to be undertaken in order to obtain public funding, price setting, reimbursement processes and inclusion of the technology in formulary lists as a requirement for funding. Such funding requirements exist in all EU member

states, although within the context of the 'subsidiarity principle'[4], health care systems differ to some extent in the mechanisms that have been set up to guide public funding of pharmaceuticals (Permanand 2006).

In Europe, the implementation of the subsidiarity principle means that the use and characteristics of funding requirements, as well as their management and implementation, fall largely, if not totally under member state competency (rather than centralized EU competency). This allows for variations in the nature of funding requirements and funding decisions between Member States (G10 Medicines Group 2002). Such variation can be attributed in part to the fact that industrial policy, public health, and health policy objectives influence the nature of pharmaceutical regulation (Permanand 2006). Industrial policy objectives focus on ensuring the productivity of the pharmaceutical industry in Europe, as well as the realisation of a single European market, which is the remit of the EU Commission. In Europe, the pharmaceutical sector employs more than 600,000 people (Kanavos et al. 2011). In contrast, public health objectives related to pharmaceutical regulation focus on ensuring public health by providing patients with access to efficacious, safe and high-quality medicines. The European Medicines Agency's (EMA) 'registrational requirement' acts as a safeguard for patient health in this respect and addresses public health objectives. Finally, health policy objectives aim to balance equity and efficiency of distribution of effective therapies for the patient population (Mossialos et al. 2004; Vogel 2004). Indeed, from a theoretical perspective, effective healthcare provision requires a balance between two potentially conflicting objectives – on the one hand i) an increase in efficiency in the production process, through the minimisation of government intervention, freely competitive markets allowing free pricing and no regulatory intervention by government bodies; and on the other hand ii) the search for an equitable distribution in society of the benefits derived from health care inputs through government intervention, such as price controls (Vogel, 2004).

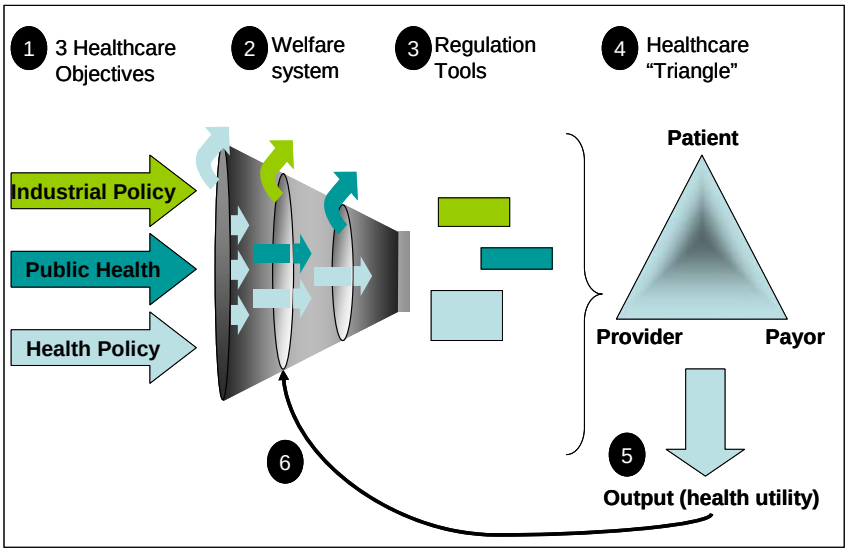
How the balance is struck between industrial policy, public health and health policy objectives varies between countries, in part due to differences in welfare systems which

[4] The notion of subsidiarity can be defined as "the principle that a central authority should have a subsidiary function, performing only those tasks which cannot be performed effectively at a more immediate or local level". (Oxford English Dictionary 2011).

reflect societal preferences. Wiktorowicz (2003) analysed the approach to pharmaceutical regulation adopted by the USA, Canada, Britain and France by considering five features for analysis – representation, process, stance, institutional power, and resources. This was coupled with an institutional framework approach to help clarify the patterns observed in the selected countries. The results of this analysis highlighted important differences in the approaches adopted towards pharmaceutical regulation, reflecting differences in how countries aim to balance objectives of timeliness of the decision-making processes along with the demanding regulatory requirements.

Figure 1.2 aims to summarise the key forces mediating the regulatory mechanisms that govern healthcare access in a particular healthcare system, and to highlight the focus of this research within the overall context. The diagram starts with the three key healthcare objectives as described above (Phase 1). The welfare system, depicted as an 'osmotic' barrier (Phase 2), favours specific types of objectives over others – thus, healthcare objectives that more closely fit into the welfare system type filter more quickly through the preference system. For example, in a member state where the pharmaceutical industry is a significant contributor to the country's gross domestic product (GDP), industrial policy objectives may carry more weight and align more closely with the welfare system. In Phase 3, regulatory mechanisms are implemented that reflect the trade-off resulting from Phases 1 and 2 (i.e. where healthcare objectives meet welfare preferences). The presence of the various types of regulatory mechanisms thus reflects the healthcare system's trade-off between objectives, within the context of the welfare system. These regulatory mechanisms then impact on the 'target' actor of the healthcare 'triangle' – be it the provider, payer or patient (or a combination of these). In this thesis, the focus will be exclusively on the payer and in particular, on how HTA bodies implement regulations and make decisions within the context of national healthcare and welfare systems – which in turn influence healthcare system outcomes (Phase 5). These outcomes are expected to either reinforce the welfare system and health care objective trade-off, or challenge it to change (Phase 6). This sixth phase, while useful to acknowledge as a key part of the context within which payers operate, is not the focus of this thesis.

**Figure 1.2 Context for analysis: the healthcare system and the influence of healthcare objectives and the welfare system on regulatory mechanisms**




Pricing and reimbursement requirements are present in most EU Member States, although their nature varies (Mossialos et al. 2004; Kanavos et al. 2011). For example, in Belgium, as in Austria, Greece, Portugal, Spain and Switzerland, price and reimbursement submissions occur as parallel but separate processes, while in the United Kingdom price approval is required, but there is currently no required reimbursement submission (Mossialos et al. 2004). In Finland, France, Italy, the Netherlands and Sweden a combined pricing and reimbursement procedure exists. Four distinct approaches to the regulation of pricing of pharmaceuticals have been identified (Mrazek 2002) – fixed pricing, cost-effectiveness pricing, profit controls and reference pricing. These approaches can be combined, and EU Member States differ in the approaches adopted (Kanavos et al. 2011). Beuscart et al. (2010) show that in France the final coverage decision actually represents a series of processes lasting several years, designed to ensure public funding is spent on technologies that bring an added benefit and that are safe and of high quality. The steps include obtaining the European marketing authorisation, control of quality by the *Agence française de sécurité sanitaire des produits de santé* (Afssaps), submission to the *Haute Autorité de santé* (HAS) for reimbursement assessment, and final negotiation with the *Comite economique des produits de santé* (CEPS) to finalise the reimbursed price (Beuscart et al. 2010).

Health Technology Assessment (HTA) is an important requirement which exists in several EU member states. HTA can be defined as a:

> The systematic evaluation of the properties, effects, and/or other impacts of health care technology (HTAi 2010)

The assessment of the evidence and the subsequent appraisal of the evidence represent two distinct stages within HTA (Sorenson 2010), and successful implementation of HTA requires multi-disciplinary assessment of the range of social, economic, clinical and healthcare system organisational consequences stemming from the introduction of a new technology relative to the existing standards of care (EUnetHTA 2010; Velasco-Garrido and Busse 2005; Henshall et al. 1997; HTAI 2007).

The nature of HTA has evolved over time - since its debut in the 1970s as a tool for controlling the use of expensive medical equipment (Jonsson and Banta 1999), HTA's remit is now broader in scope and includes pharmaceutical technologies in its assessment programme. While originating in the USA, HTA is now established in many EU Member States (Jonsson 2002; Banta and Oortwijn 2000a, 2000b; Gulacsi 2001; Oliver et al. 2004; Velasco-Garrido and Busse 2005; Sorenson 2010). Economic evaluations are becoming an increasingly wide-spread requirement across European countries (Kanavos et al. 2000;Nuijten et al. 2001; Drummond 2003; Nuijten and Kosa 2004, Kanavos et al. 2011). The nature of HTA varies between European countries both in terms of the assessment of the evidence and the appraisal of the evidence, as well as in the role of HTA within each healthcare system (Jonsson 2002; Banta and Oortwijn 2000; Gulacsi 2001; Oliver et al. 2004; Velasco-Garrido and Busse 2005; Sorenson 2010).

The National Institute for Health and Clinical Excellence (NICE) in England and Wales, the Scottish Medicines Consortium (SMC), the Dutch College voor Zorgverzekeringen (CVZ), and the French Haute Autorité de Sante (HAS) are examples of HTA bodies in Europe and have been chosen as the focus of this research. The implications of HTA decisions are further explored in the section that follows.

## 1.3 Implications of HTA decisions for patients, providers, manufacturers and health policy makers

> "[In France] Between 1990 and 2009, public prices of reimbursed medicines decreased by 20.6%, although in the same time period, inflation increased by 38.4%"5 (Les Entreprises du Medicament 2010 p1)

> "Scottish cancer patients should be able to access the same standards of cancer care as their English counterparts." (RCF Chief Executive Andrew Wilson in Herald Scotland 2010 p1)

This thesis distinguishes between coverage of pharmaceuticals – which reflects the willingness to fund a particular technology that is deemed to be of value – and reimbursement of pharmaceuticals, which is the implementation of the coverage decision in regulatory mechanisms that allows for their inclusion/exclusion from a reimbursement list and from use within a healthcare system. HTA bodies vary in the nature of recommendations they make, the evidence and processes they use to make such recommendations, and whether their recommendations have an advisory or regulatory role (Sorenson 2010, Sorenson et al 2008). For example, the CVZ will concurrently examine the degree of value of intervention and whether it should be funded by the healthcare system (i.e. coverage), but also the price of the intervention and whether the technology should be clustered, independently priced, or placed on an 'expensive drug' list (i.e. reimbursement). In France, on the other hand, the HAS, and specifically the Comitee de Transparence, issues advice on the value of the technology for a particular population (ASMR) and advises on the patient population eligible for treatment (i.e. coverage), while the CEPS negotiates price and volume agreements with manufacturers on the basis of the ASMR rating provided by the HAS (i.e. reimbursement) (Sorenson 2010)[6]. Technologies with an ASMR rating of V are those that represent no incremental value and can only be considered for reimbursement if its cost is below that of its comparator[7], highlighting that there is no willingness to fund

---

[5] Original quotation: "Entre 1990 et 2009, les prix publics des médicaments remboursables ont diminué de 20,6% alors que dans le même temps, l'inflation augmentait de 38,4%"

[6] For technologies with an ASMR I-III rating the healthcare system is willing to pay more (premium) over standard of care as they are associated with significant incremental benefit, while technologies with an ASMR IV rating can expect to receive the same level of funding as that obtained by their comparator technology.

[7] It is unclear however, to what extent manufacturers in France who receive an ASMR V rating opt for a reduction in price to obtain reimbursement or if they withdraw from the healthcare system.

such technologies without cost savings. The decision output by the HAS diverges from NICE, SMC and the CVZ which consider both coverage and reimbursement in their decision-making, and in which non-recommendations preclude the inclusion on reimbursement lists, or highly discourage the use of the technology. With the considerable intricacies and heterogeneity in the reimbursement mechanisms and pharmaceutical pricing policies in place, this thesis aims to focus on assessing the factors driving the willingness to fund a particular technology, rather than the reimbursement mechanisms and negotiations that follow.

Coverage decisions represent a key point within the complex decision-making process that governs funding and access for pharmaceuticals. Coverage decisions are of interest because of their implications for patients, providers, manufacturers and health policy makers.

There is particular interest from multiple stakeholders in HTA because of the implications that coverage decisions can have on access to pharmaceuticals, and on health outcomes. From an efficiency-driven approach, access to health care can be justified from an economic perspective because of its presumed benefits in improving the health of the population (both from a clinical point of view but also from a productivity point of view) (Gulliford and Morgan 2003). However, such an approach may disregard the needs of vulnerable populations (e.g. the elderly, those with mental health problems, prisoners etc.). An assessment of the level of access to healthcare necessitates consideration of various components including the comprehensiveness of the services offered, when and to whom these are offered, and how such services impact on health outcomes (Gulliford and Morgan 2003). Two facets of 'access' become apparent in the literature and are of relevance in any analysis of access. From one perspective, access can be defined as equal utilisation of healthcare resources (Donabedian 1972 in Gulliford and Morgan 2003). Alternatively, access can be defined as the availability of a service, where access is about equal opportunity, not equal utilisation (Mooney 1983). National HTA bodies fit within the latter definition of access, in that they have the ability to define the availability of a therapy, by determining whether or not a pharmaceutical technology will be publicly funded, and if so, to what degree.

The implications of coverage decisions for patients and providers are more visible when HTA bodies decide not to reimburse a product for a particular patient population, as highlighted in the following excerpt: "…[Cancer] medicines are licensed and working yet we in Britain aren't allowed to access them. While we wait for NICE to decide, patients are dying" (Kate Spall on BBC Breakfast News, May 13th 2008). Thus, non-recommendations from HTA bodies can be perceived as decisions that reduce treatment options for particular patient groups. The excerpt also highlights the potential implications linked with coverage decisions for health outcomes, in this example, death.

While recognising the role of broader medical and non-medical factors on health outcomes such as the life expectancy of the population, the literature does provide evidence that links access to pharmaceuticals with better health outcomes. There are several studies reported in the literature that assess the impact of pharmaceutical access, as measured by levels of pharmaceutical expenditures, on life expectancy. Within OECD countries, Miller and Frech (2000) calculated the ratio of money spent on pharmaceuticals versus the life expectancy gain and concluded that pharmaceutical consumption was associated with a statistically significant positive effect on life expectancy. In a further study by Frech et al. (2004), the authors confirmed that a strong statistical effect was observed between pharmaceutical consumption and life expectancy at both 40 and 60 years.

In addition, they also examined the impact of pharmaceutical consumption on quality of life, using disability adjusted life expectancy (DALE) as the measure of quality of life. In their analysis they found that pharmaceutical use was associated with an even larger effect than that observed on life expectancy: positive statistically significant effects of pharmaceutical utilisation on DALE were observed both at birth and at 60 years. In examining the relationship between pharmaceutical expenditure and life expectancy, Caliskan (2009) took the methodology one step further by adjusting for other factors known to have an impact on life expectancy, including socio-economic status, lifestyle factors and demographic factors. The results suggest that pharmaceutical expenditure impacts positively on life expectancy, and further note that this effect varies by age and gender. A similar study examining the relationship between pharmaceutical expenditure and life expectancy in Canada also showed that the association between pharmaceutical expenditure and health outcomes was statistically significant (Cremieux

et al. 2005). In addition, the authors examined variations in pharmaceutical expenditure among the Canadian provinces and demonstrated that provinces with higher pharmaceutical spending were those provinces with better health outcomes (as measured by infant mortality and life expectancy, among other endpoints). The authors estimated that if all provinces were to increase pharmaceutical spending to the level of the 'high spending' provinces, this could lead to an additional 6 months of life expectancy at birth and an estimated 584 fewer infant deaths annually and (Cremieux et al. 2005).

In parallel to studies on the impact of pharmaceutical access on life expectancy and infant mortality, a study in Germany examined the role of pharmaceutical interventions on the decline in cardiovascular-related deaths (Häussler et al. 2007). The analysis involved examining specific types of pharmaceutical interventions, alongside other medical interventions, to assess to what degree the introduction of pharmaceutical interventions impacted on mortality from cardiovascular causes. The results of the regression analysis found that the different types of pharmaceutical interventions tested (from use of ACE inhibitors and channel blockers, to the use of diuretics and beta-blockers) had individually and collectively statistically significant effects on the decline of mortality.

Coverage decisions not only have implications for patients and providers but also for the pharmaceutical market and health-policy makers. Within the European internal market, pharmaceuticals are categorised like any other good, meaning that free movement and access to the market is advocated and required as a condition for participation in the market. Thus, the availability of pharmaceuticals and the time taken for them to be available within the market is not only a concern for patients and health care providers, but also for national and supra-national government entities such as the European Commission. To this end, the EU Transparency Directive 89/105/EEC(The Council of the European Communities 1989) adopted in 1989 requires EU Member States to implement "objective and verifiable criteria" in determining the price for novel pharmaceutical compounds and in determining whether such a compound should be funded/reimbursed (The Council of the European Communities, 1989 p. 67). In addition, the EU Directive sets a time limit of 180 days for the completion of the

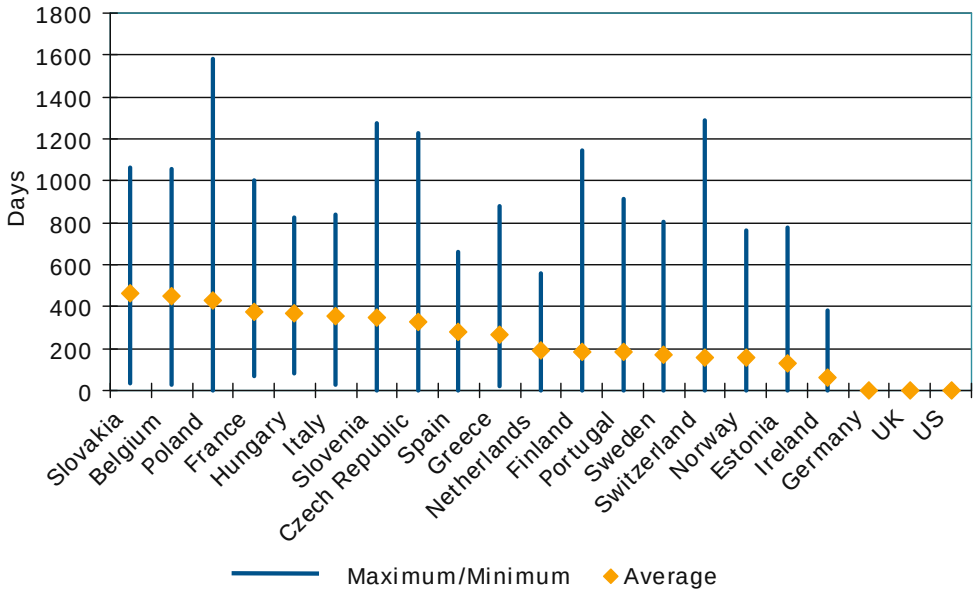
coverage decision-making process, which Member States are requested to implement (unless a 'clock-stop' is required due to the need for further data/clarification) [8].

Despite the existence of the above directive for more than 20 years, and the presence of a centralised procedure for granting marketing authorisation since 1993 (EEC No. 2309/93 (The Council of the European Communities 1993), variation between EU Member States in the availability of medications continues to be documented in the literature (Anis et al. 2001; Wilking and Jonsson 2005; IMS 2007; Jonsson and Wilking 2007). Wilking and Jonsson (2005), and Jonsson and Wilking (2007) provide evidence of variation in the diffusion of new medications for cancer illnesses in Europe, which suggests that access to new cancer therapies is to a large extent dependent on a patient's country of residence. Specifically, the authors venture to suggest (but not test) processes that augment variation in access to care, including the drug approval process, the role of health economics in decision making, and budgetary issues. In Jonsson and Wilking (2007), these factors are re-examined over a larger sample size (both in terms of drugs and countries), but data on the impact of factors such as reimbursement processes and cost-effectiveness requirements is still lacking.

As documented by IMS data collected between 2002 and 2006, there appears to be variation between EU Member States in the number of novel medications accessible to patients, and also variation in the time taken by the decision-making authorities to grant access to medications (Figure 1.3). In the majority of cases, the time between marketing authorisation and a final coverage decision was often higher than the 180 days stipulated in the EU Transparency Directive. This analysis was recently repeated and confirmed earlier results, highlighting that significant variation remains between EU Member States in the percentage of licensed pharmaceuticals available for prescription (39% to 86%), and in the time taken to complete pricing and reimbursement processes (88 to 392 days) (EFPIA 2010).

[8] It should be noted that the European Commission has recently launched a consultation on the EU Transparency Directive, with the aim to modernize the directive to reflect the changes in the EU pharmaceutical market over the past 20 years.
http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/370&format=HTML&aged=0&language=EN&guiLanguage=en

**Figure 1.3** Range of time (days) between EMA marketing authorisation and market access (hospital and retail combined)




*Source:* IMS 2007.

*Note*: In this IMS analysis, time to market access refers to the time from marketing authorisation to the availability of the technology funded by the particular healthcare system. For the UK, while technologies can be prescribed within the NHS shortly after marketing authorisation, the market access of the technology is limited until it has completed HTA procedures within the UK.

Variation in the availability of medications and the average time to access them could be influenced by a number of factors, including the nature of the decision-making procedures, but also the manufacturer's submission strategies (for example, launching in free-pricing countries first), product characteristics (therapeutic value), country GDP, and the objectives of the welfare and health care systems. The need for Member States to work harder to adhere to the Transparency Directive is highlighted in 2002 in the G10 Report, which recommends that "… Member States should examine the scope for improving time taken between the granting of a marketing authorisation and pricing and reimbursement decisions in full consistency with Community Legislation" (G10 Medicines Group 2002 p. 13).

Thus, coverage decisions are of interest because of their far reaching consequences for patients, health care providers, manufacturers and health policy makers. HTA bodies face distinct pressures from each of these stakeholders when making their decisions. Therefore, it is important to understand what factors are in fact driving patterns of coverage decisions within and between HTA agencies.

## 1.4 Research Question

This introductory chapter has aimed to provide the context and rationale for the research question of this thesis. Across EU Member States, the pharmaceutical market is heavily regulated, and public funding of pharmaceuticals is guided by specific processes and regulations, which vary across countries. Such variation across EU Member States is in large part due to diverse approaches reflecting differences in national healthcare and welfare objectives. This is accompanied by an observable variation in the level of access to medication between EU Member States. European Member States implement regulations aiming to control the cost of supplying medications to the healthcare system by deciding what technology should or should not be publicly funded, and if it is, under what conditions (its price, who should have access to it, and who should prescribe it). HTA is a specific type of assessment required in many European countries to obtain public coverage of pharmaceutical technologies. Within this context, HTA bodies have a role in defining whether public funding should be allocated to the technology. The interest in analysing such coverage decisions is heightened by the fact that ultimately the decisions may not only impact on medication price and volume, but are also likely to impact on patients' access to medications, as well as on the behaviour of prescribers, manufacturers and policy makers. Moreover, the need for difficult decision-making with regard to healthcare spending and allocation of resources is amplified as demand for healthcare increases while at the same time governments and healthcare providers struggle to manage healthcare expenditure, giving rise to tensions between HTA agencies, the public, healthcare providers, manufacturers and health policy makers. In this context of difficult decision-making, there is a recognised need for the transparent communication of the processes and criteria adopted by HTA bodies in making decisions about public funding and granting (or not) of access to pharmaceuticals. This thesis analyses the factors that drive HTA coverage decisions and public funding of pharmaceuticals in a selection of European countries, namely UK, France, and the Netherlands.

Specifically, this thesis aims to:

- Describe the range of factors taken into account by NICE, SMC, CVZ and HAS in their decision making, including the clinical/economic evidence considered, the process

through which technologies were appraised and the socio-economic context in which the appraisals took place.

- Explore the impact of such factors on coverage decisions. Given the multidisciplinary nature of HTA decision-making and the high degree of stakeholder involvement in and around HTA decision-making, it is hypothesised that HTA coverage decisions are influenced not only by the evidence supporting the technology, but also by the assessment processes used and the context in which they operate. This will be tested by identifying, for each HTA body, the set of explanatory variables that significantly impact on its decision outcomes, while adjusting for the presence of other confounding factors. Decision outcomes will be defined as the log odds of restriction versus recommendation or non-recommendation versus recommendation.
- Assess whether HTA bodies (NICE, SMC, CVZ, HAS) differ in terms of the coverage decisions they make. This hypothesis will be tested in a pooled analysis in which the HTA effect will be assessed while adjusting for a range of confounders.

In order to address the research question, the thesis is structured into nine chapters. Following this introductory chapter, the second chapter identifies gaps in the literature that this thesis seeks to address, and provides the analytical framework for the research. The methods chapter (Chapter 3) identifies the sets of factors that, through the literature review, have been identified as potential determinants of coverage decisions, and the methods by which these factors will be considered in single HTA and pooled analyses. Chapters 4 to 7 present the quantitative analyses performed to assess the factors driving coverage decisions individually within each of the four HTA bodies included in the research. Chapter 8 examines the factors driving coverage decisions across the pooled sample of appraisals from the four HTA bodies and Chapter 9 presents the conclusions and key policy implications.

## 1.5 References

Abel-Smith, Brian, and P Grandjeat. 1978. *Pharmaceutical consumption: trends in expenditure : main measures taken and underlying objectives of public intervention in this field.* Luxembourg: Office for Official Publications of the European Communities.

Anis, AH, D Guh, and X Wang. 2001. A dog's breakfast: prescription drug coverage varies widely across Canada. *Med Care* 39 (4):312-4.

Banta, D., and W. Oortwijn. 2000a. Health technology assessment and health care in the European Union. *Int J Technol Assess Health Care* 16 (2):626-35.

———. 2000b. Introduction: health technology assessment and the European Union. *Int J Technol Assess Health Care* 16 (2):299-302.

Beuscart, R., Chazardb, E., and Soufa, N. 2010. De l'innovation au remboursement / From innovation to reimbursement. *IRBM* 31 (1): 26-29.

Barr, NA. 2001. *The welfare state as piggy bank: information, risk, uncertainty, and the role of the state*: Oxford University Press.

Caliskan, Z. 2009. The relationship between pharmaceutical expenditure and life expectancy: evidence from 21 OECD countries. *Applied Economics Letters* 16: 1651-1655

Cremieux, P. Y., M. C. Meilleur, P. Ouellette, P. Petit, M. Zelder, and K. Potvin. 2005. Public and private pharmaceutical spending as determinants of health outcomes in Canada. *Health Econ* 14 (2):107-16.

Drummond, MF. 2003. The use of health economic information by reimbursement authorities. *Rheumatology* 42 (Suppl. 3):iii60-iii63.

Dukes, M, ed. 2003. *Drugs and money : prices, affordability and cost containment* Amsterdam: IOS Press Ohmsha.

EUnetHTA. 2011. "HTA definition". http://www.eunethta.net/Public/About_ EUnetHTA/HTA/. Viewed on 2nd February 2011.

European Commission. 2009. Pharmaceutical Sector Inquiry – Final Report. Adoption Date: 8 July 2009. http://ec.europa.eu/competition/sectors/ pharmaceuticals/ inquiry/staff_working_paper_part1.pdf

European Federation of Pharmaceutical Industries and Associations (EFPIA). 2010. Patients W.A.I.T. indicator – 2010 Report. http://www.efpia.eu/Content/ Default.asp?PageID=559&DocID=10200.

Frech, A., Miller, R.D. 2004. The Effects of Pharmaceutical Consumption and Obesity on the Quality of Life in the Organization of Economic Cooperation and Development (OECD) Countries. *PharmacoEconomics* 22(2): 25-36

G10 Medicines Group. 2002. High Level Group on innovation and provision of medicines; recommendation for action. Belgium: European Communities.

Garrett, M. D., M. I. Walton, E. McDonald, I. Judson, and P. Workman. 2003. The contemporary drug development process: advances and challenges in preclinical and clinical development. *Prog Cell Cycle Res* 5:145-58.

Gulacsi, L. 2001. Health technology assessment in Central and Eastern Europe. *Eurohealth* 7:34–36.

Gulliford, M., and Morgan, M., eds. 2003. *Access to Health Care*. New York: Routledge.

Häussler, B., Schiffhorst, G., Gothe, H. and Hempel, E. 2007. The impact of pharmaceuticals on the decline of cardiovascular mortality in Germany. *Pharmacoepidemiology and Drug Safety* 16: 1167–1176.

Henshall C., W. Oortwijn, A. Stevens, A. Granados, and D. Banta. 1997. Priority-setting for health technology assessment. Theoretical considerations and practical approaches. Priority-setting subgroup of the EUR-ASSESS Project. International *Journal of Technology Assessment in Health Care* 13:144-185.

HTAI. 2007. Health Technology Assessment International. http://www.inahta.org/upload/HTA_resources/AboutHTA_Policy_brief_on_HTA_Introduction_to_Objectives_Role_of_Evidence_Structure_in_Europe.pdf. Viewed on 2 June 2009.

IMS. 2007. Patients W.A.I.T. Indicator - Phase 7 Report: IMS.

Jacobzone, S. 2000. *Pharmaceutical policies in OECD countries: reconciling social and industrial goals*: Organisation for Economic Co-operation and Development.

Jonsson, E. 2002. Development of health technology assessment in Europe - a personal perspective. *International Journal of Technology Assessment in Health Care* 18 (2):171-183.

Jonsson, E., and D. Banta. 1999. Treatments that fail to prove their worth. Interview by Judy Jones. *BMJ* 319 (7220):1293.

Jonsson, B., and Wilking,N. 2007. A global comparison regarding patient access to cancer drugs. *Annals of Oncology* 18: iii1-iii78.

Kanavos, P, P Trueman, and A Bosilevac. 2000. Can economic evaluation guidelines improve efficiency in resource allocation? The cases of Portugal, the Netherlands, Finland, and the United Kingdom. *International Journal of Technology Assessment in Health Care* 16 (4):1179-1192.

Kanavos, P., S. Vandoros, R. Irwin, E. Nicod, and M. Casson. 2011. Differences in costs of and access to pharmaceutical products in the EU. Brussels, European Parliament, 2011. http://www.europarl.europa.eu/activities/committees/studies/download.do?language=en&file=35108

Les Entreprises du Medicament (LEEM). 2010. Évolution des prix des médicaments en France. http://www.leem.org/medicament/evolution-des-prix-des-medicaments-en-france-408.htm. Viewed on 5 January 2011

Maynard, A. and Bloor, K. 2003. Dilemmas In Regulation of the Market for Pharmaceuticals. *Health Affairs* 22: 31-41.

McGuire, A., M. Drummond, and F. Rutten. 2004. "Reimbursement of Pharmaceuticals in the European Union" in E. Mossialos, M. Mrazek and T. Walley (Eds) *Regulating Pharmaceuticals in Europe. Striving for Efficiency, Equity and Quality*. United Kingdom: Open University Press.

Miller, RD, and HE Frech. 2000. Is there a link between pharmaceutical consumption and improved health in OECD countries? *Pharmacoeconomics 2000* 18 (Suppl. 1):33-45.

Mooney, G. H. 1983. Equity in health care: confronting the confusion. *Eff Health Care* 1 (4):179-85.

Mossialos, E., Mrazek, M., and Walley, T. 2004. *Regulating pharmaceuticals in Europe: striving for efficiency, equity and quality*. Maidenhead, UK: Open University Press.

Mrazek, M. 2002 Comparative Approaches to Pharmaceutical Price Regulation in the European Union. *Public Health* 43: 453-461.

Nuijten, MJC, P Berto, G Berdeaux, J Hutton, F.-U Fricke, and F.A. Villar. 2001. Trends in Decision-Making Process for Pharmaceuticals in Western European Countries: A Focus on Emerging Hurdles for Obtaining Reimbursement and a Price. *The European Journal of Health Economics* 2 (4):162-169.

Nuijten, M. J., and J. Kosa. 2004. Pricing of pharmaceuticals. Assessing the pricing potential by a pricing matrix model. *Eur J Health Econ* 5 (2):110-5.

Oxford English Dictionary. 2011. Second edition. Online version March 2011. http://www.oed.com:80/Entry/193007. Viewed on 05 April 2011.

Oliver, A, E Mossialos, and R Robinson. 2004. Health technology assessment and its influence on health-care priority setting. *International Journal of Technology Assessment in Health Care* 20 (1):1-10.

Permanand, G. 2006. *EU pharmaceutical regulation: the politics of policy-making.* Manchester: Manchester University Press

PhRMA. 2007. Innovation. www.phrma.org/innovation/. Viewed on 16 February 2009.

Reinhardt, U.E. 1990. Economic relationships in health care, in OECD Health Care Systems in Transition: The Search for Efficiency. Paris: Organisation for Economic Co-operation and Development.

Rice, T., Biles, B., Brown, E., Diderichsen, F., and Kuehn, H. 2000. Reconsidering the Role of Competition in Health Care Markets: Introduction. *Journal of Health Politics, Policy and Law* 25: 863-873.

Saltman, R., Figueras, J., and Sakellarides, C. (Eds). 1998. Critical Challenges for Health Care Reform in Europe. Buckingham, UK: Open University Press.

Sorenson, C. 2010. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison. Issue in International Health Policy Volume 91. USA: The Common Wealth Fund.

Taylor, R S, M F Drummond, G Salkeld, and S D Sullivan. 2004. Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. *BMJ* 329 (7472):972-975.

The Council of the European Communities. 1989. Council Directive 89/105/EEC

———. 1993. Council of Regulation (EEC) No 2309/93.

Velasco Garrido, M., and Busse, R. 2005. Health Technology Assessment—An Introduction on Objectives, Role of Evidence, and Structure in Europe. Policy Brief. Brussels, European Observatory on Health Systems and Policies.

Vogel, R. J. 2004. Pharmaceutical pricing, price controls, and their effects on pharmaceutical sales and research and development expenditures in the European Union. *Clin Ther* 26 (8):1327-40; discussion 1326.

Wilking, N, and B Jonsson. 2005. A pan-European comparison regarding patient access to cancer drugs. Stockholm: Karolinksa Institutet and Stockholm School of Economics.

Wiktorowicz, ME. 2003. Emergent Patterns in the Regulation of Pharmaceuticals: Institutions and Interests in the United States, Canada, Britain, and France. *Journal of Health Politics, Policy and Law* 28 (4):615-658.

# 2 Framing the Research Question

In order to frame the research question, an introductory example is presented, which summarises the coverage decisions by two HTA agencies, the Scottish Medicines Consortium (SMC) in Scotland and the Transparency Commission of the Haute Autorité de Santé (HAS) in France. In 2007, both the HAS and the SMC reviewed a request for reimbursement of dasatinib, a new therapy for the treatment of chronic and advanced phases of chronic myeloid leukaemia (CML). The excerpts below summarise the advice given by both HTA bodies:

> Scottish Medicines Consortium (SMC) Drug advice on dasatinib:
>
> "dasatinib [...] is accepted for restricted use within NHS Scotland for the treatment of adults with chronic myeloid leukaemia (CML) with resistance or intolerance to prior therapy including imatinib mesylate.
> It should be restricted to use in patients who are in the chronic phase of the disease. The manufacturer's justification of the treatment's cost in relation to its health benefits for the accelerated or blast phases was not sufficient to gain acceptance by SMC" (SMC 370/07 2007 p.1)

> Commission de la Transparence (Haute Autorité de Santé) advice on dasatinib[1][2]:
>
> "In chronic phase chronic myeloid leukemia (CML), after resistance or intolerance to previous therapy including imatinib, Sprycel brings an ASMR of level II (important) compared to current treatment. In accelerated or blast phase chronic myeloid leukemia after resistance or intolerance to imatinib and in acute Ph+ lymphoblastic leukemia after resistance or intolerance to previous therapy, Sprycel brings and ASMR of level I (major) compared to current treatment" (Commission de la Transparence 2007 p. 11)

[1] "Dans la LMC en phase chronique, après résistance ou intolérance à une thérapie antérieure incluant imatinib, Sprycel apporte une ASMR de niveau II (importante) par rapport à la prise en charge thérapeutique actuelle. Dans la LMC en phase accélérée ou blastique après résistance ou intolérance à l'imatinib et dans la leucémie aiguë lymphoblastique Ph+ après résistance ou intolérance à une thérapie antérieure, Sprycel apporte une ASMR de niveau I (majeure) par rapport à la prise en charge thérapeutique actuelle.".

[2] Note: Sprycel© is the tradename for dasatinib.

As highlighted by the quoted excerpts, both HTA agencies come to a similar decision with regard to the public funding of dasatinib for chronic phase CML, but not for advanced stages of CML. The SMC and HAS recommended dasatinib for the treatment of *chronic* phase CML. The SMC states that dasatinib is "*accepted for restricted use"* while the Transparency Commission grants it an ASMR of level II. The "amélioration du service médical rendu" (ASMR) is a rating which aims to capture the extent to which a medication can improve the outcomes for patients and address significant unmet medical need. A rating of 'I' is given to those medicines considered to bring a significant medical improvement, versus a rating of 'V' for those medicines considered to provide no improvement. Thus, by giving an ASMR rating of 'II', the Transparency Commission recognises that dasatinib brings an important medical improvement versus the standard of care (imatinib) in the treatment of chronic phase CML. However, opposing decisions were made with regard to the recommendation of dasatinib for treatment in the advanced phases of the disease. In Scotland, the SMC did *not* recommend use of dasatinib in advanced phases of CML. In contrast, the French Transparency Commission awarded dasatinib an ASMR rating 'level I' for treatment of advanced phase CML for providing a major medical improvement to patients.

Within this particular case-study, what are the factors that can explain the coverage decisions made in France and in Scotland with regard to dasatinib? Was the same evidence (clinical and economic) considered by both HTA bodies? Are the decision-making processes different and could that explain their recommendations? Perhaps the healthcare system, welfare system and societal context influenced the HTA agencies' decision? Is a difference in recommendations made by these two bodies consistently observed, or is this a one-off difference? This example sets the scene for the focus of this research project – that is, the in-depth analysis of what factors drive HTA coverage decisions.

This chapter aims to frame the research question by combining theoretical and empirical elements to develop an analytical framework that can be used to analyse coverage decisions and the factors driving these decisions. The chapter is divided into four sections. The first section provides an overview of theoretical approaches to understanding HTA decision-making. The literature relevant to the research question is then presented and analysed, focusing on the evidence, process and context factors that

have been associated with coverage decisions. The third section highlights the gaps and limitations of the currently available literature, and identifies those gaps which the thesis will attempt to address. The final section brings together the theoretical concepts and empirical data to propose an analytical framework that will help shape the methods adopted to address the research question.

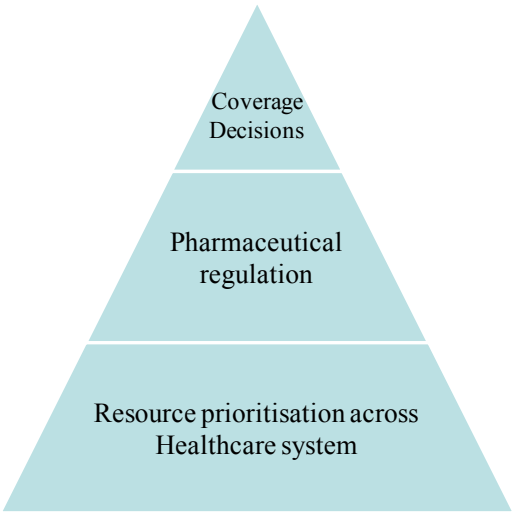
## 2.1 Theoretical models for understanding coverage decision-making

The need to prioritise resources in a healthcare system is linked to two phenomena: firstly, the demand for health and healthcare by the population and secondly, the recognition that the resources available to meet those demands (i.e. healthcare) are limited. This has put pressure on healthcare systems to develop ways to prioritise the use of healthcare resources. Indeed, from a theoretical perspective, priority-setting within the healthcare system requires an equilibrium between two potentially conflicting objectives – on the one hand (i) the objective of increasing the efficiency of the production of health gains for the population, through freely competitive markets allowing free pricing and no regulatory intervention by government bodies; and on the other hand (ii) the objective of achieving an equitable distribution in society of the benefits derived from health care resources through government intervention (Vogel, 2004). HTA decision-making on whether to recommend a new medication for public funding is a specific example of prioritisation of resources within a healthcare system. Figure 2.1 (below) outlines how coverage decisions fit within the concept of priority setting. The complexity around evaluating priority-setting decisions and their success is linked to the fact that various criteria can be used in prioritizing healthcare resources, such as: ethical considerations, efficiency considerations and effectiveness considerations (Sibbald 2009; Al et al. 2004; Permanand 2006). If a healthcare system objective is to strive for effectiveness over efficiency, the prioritisation exercise would yield different results as supposed to a healthcare system that aims to sacrifice some efficiency gains to give more weight to ethical considerations, for example.

In Europe, when making funding decisions, HTA bodies can be influenced by potentially conflicting objectives: public health objectives, health policy objectives and industrial policy objectives (Permanand 2006). Industrial policy objectives focus on ensuring the productivity of the pharmaceutical industry in Europe, as well as the realisation of a single European market, which is the scope of the EU Commission

(Directorate General for Enterprise). Public health objectives related to pharmaceutical regulation focus on ensuring public health by providing patients with access to efficacious, safe and high quality medicines. The European Medicines Agency's (EMA) 'registrational requirement' acts as a safeguard for patient health in this respect and addresses Public Health objectives. Finally, health policy objectives aim to balance equity and the efficiency of distribution of effective therapies for the patient population (Mossialos et al. 2004; Vogel 2004).

**Figure 2.1 HTA Coverage Decisions as a mechanism of prioritisation of resources within the healthcare system**




Due to the complexity surrounding priority-setting decisions, a number of different suggestions have been made regarding how to make decision-making more systematic. An important development recorded in the literature over the last few decades is the move towards evidence-based policy and evidence-based medicine as a means of making prioritisation decisions more rational, systematic, justifiable and transparent (Daniels and Sabin 1997; Drummond, Schwartz et al. 2008). Within the sphere of coverage decision-making, the HTA movement can be linked to this phenomenon, developed as a means of systematically evaluating evidence to come to a conclusion about the utilisation and funding of a new health technology.

Theoretical models have been developed to characterise how evidence has or can be used in policy-making. In social policy, several theoretical models have been proposed to explain how and to what extent evidence is incorporated in policy-making (Weiss 1979). For this research question, two particular models proposed by Weiss are

relevant: (i) the 'problem-solving model' because it represents what evidence-based funding strives for, and (ii) the 'interactive model' which represents the situation where decisions are influenced by diverse factors, many of which are not evidence-based.

The premise of the 'problem-solving model' is that evidence, especially developed to answer a specific policy problem, is directly applied to solve the policy decision problem, in this case, coverage decisions. As outlined by Weiss (1979), the 'problem-solving model' suggests that evidence is integrated in decision-making through the following steps: "… definition of pending decision → identification of missing knowledge → acquisition of social science research → interpretation of the research for the decision context → policy choice" (Weiss 1979 p. 428). Thus, in this model, the evidence is tailor-made to answer a specific policy-maker's question and assumes that evidence is the primary influencer of the decision. It also assumes that, by directly influencing the decision, the evidence will also lead to a rationalization of resource use in accordance with the evidence. Furthermore, one of the key conditions of this model is that there is alignment between the researcher and decision-maker in terms of the problem definition and objectives.

While this model was originally developed by Weiss (1979) to characterize the use of social research evidence, it aptly describes the ideal world of evidence-based policy making, where relevant, high-quality evidence reduces uncertainty in HTA decision-making about the introduction of a new therapy in the healthcare system, and is thus the key driver of the decision. Clinical and economic evidence are key drivers of HTA coverage decision-making, as they are a distinct and important part of the assessment and appraisal process. In addition, in England, NICE commissions specific third-party assessments of economic evidence specifically for the scope of the appraisal being undertaken (Williams and Bryan, 2007). The CVZ in the Netherlands, with its 'conditional reimbursement scheme', provides another example of a 'problem-solving model'; i.e. reimbursement status is granted over a three-year period, on the condition that effectiveness and cost-effectiveness data are collected, as per a mutually agreed protocol, to confirm the reimbursement status of the medication with real-life data at a defined time point in the future.

There are several challenges however, when it comes to the implementation of the 'problem-solving model' in HTA decision-making. First of all, the model assumes that all evidence consistently supports a specific decision which is preferable to all other possible decisions (Weiss 1979). However, in reality, different sources of evidence can conflict with each other, and the need to combine different sources of evidence involves subjective value judgments about which aspects of the evidence are of greater or lesser relevance to the decision-problem.

> "The problem that remains is: priority setting involves the adjudication between many relevant values and that people (and disciplines) will disagree about which values should dominate in any specific priority setting context and there is no agreed upon normative approach for resolving the disagreement" (Sibbald et al. 2009 p. 10).

This suggests it is important to distinguish between accessibility to high-quality, relevant evidence which this model emphasizes, and the second step which involves the acceptability of the evidence, which is linked to the value judgements that surround the decision-problem, as highlighted by Sibbald et al. (2009) and Williams and Bryan (2007).

Another implementation challenge associated with the 'problem-solving model' is that it assumes that there are only two stakeholders, namely the decision-maker and the researcher. However, in reality, in many circumstances this is not the case. For example, NICE implements a broad and complex consultation programme for each of its appraisals to ensure that a variety of stakeholders are consulted and involved in the appraisal process. Stakeholders in the NICE process include clinicians, patient groups, carer groups, manufacturers. Stakeholders involved in the SMC process are equally varied. One of the key stakeholders for the HAS is the CEPs, the committee which uses HAS output to negotiate a reimbursed price. The 'problem-solving' model does not factor in the role of these additional stakeholders in the coverage decision-making process.

An additional hurdle for the applicability of this model to coverage decisions includes the fact that the model assumes that evidence reduces the level of uncertainty around a decision-problem. However, the degree of uncertainty remaining around the decision to fund new medication is often high, despite the evidence available. For example, the appraisal of adalimumab for the treatment of ankylosing spondylitis by NICE (2008),

included the assessment of five different cost-utility models that pointed to different base case incremental cost-effectiveness results. Thus, despite a significant volume of economic evidence, the uncertainty for the Appraisal Committee remained – to the extent that the final appraisal allocated a range cost-effectiveness ratio for adalimumab, rather than a specific number, due to the lack of consensus.

A final challenge to the implementation of the 'problem-solving' model is that it assumes aligned objectives between the evidence provider and the decision-maker. However, there are empirical examples where objectives are not shared by both parties, and where objectives are unclear. For example, Al et al. (2004) describe four HTA decision-makers working at different decision-nodes within the system that had different objectives when making prioritisation decisions (i.e. equity considerations, efficiency considerations etc.). The authors stress the need to clarify objectives between HTA and the evidence provider so that the evidence is helpful in solving the decision-problem. The authors also show that there are implicit HTA preferences that influence their objectives, which correlates well with qualitiative findings from other research done on HTA preferences (Anell and Persson 2005; Haslé-Pham et al. 2005; Sinclair et al. 2008; Vuorenkoski et al. 2008). Thus, the objectives of HTA decision-making are not always transparent and known to the researcher or healthcare system.

As an alternative to the 'problem-solving model' and in recognition of the complex web of factors (both explicit and implicit) that influence coverage decisions, the 'interactive model' is presented here as a model that aims to capture the situation in which the evidence used by the decision-maker is taken from a multitude of sources (Weiss 1979). As Weiss (1979 p. 429) explains, "In this model, the use of research is only one part of a complicated process that also uses experience, political insight, pressure, social technologies, and judgment". Important components of decision-making within this particular model include the need for negotiation between different perspectives and objectives, the need to align various points of view and a need to come to a solution. While evidence does play a role, it is not the key influencer of the decision. In this model, the responsibility is placed is on the decision-maker to gather what evidence there is from numerous sources – there is no specific evidence commissioned to address a specific decision-problem.

When we consider this model and its applicability to HTA decision-making, evidence from the literature does suggest that, apart from the evidence provided by clinical and economic evaluations, HTA agencies consider several other factors when making their decisions (Ross 1995; OECD 2005; Drummond, Brown et al. 2003; Brouselle and Lessard 2011; Iversen and Vondeling 2007; Al et al. 2004). This model would also reflect the fact that HTA agencies must juggle between addressing healthcare policy, public health, and industrial policy objectives (Permanand 2006). This model, however, goes against the principle that policy-making, and in particular HTA, should be primarily evidence-based. In addition, the 'interactive model' assumes that no bespoke research is done to address decision-maker evidence needs. In reality, HTA bodies can commission specific research, and manufacturers themselves commission specific research to address the HTA body's questions as specified within their methodology guidelines. For example, NICE's Multi Technology Appraisal process includes the commissioning of specific research for the appraisal. Thus, the 'interactive model' also faces implementation challenges and represents, to some extent, an opposing view of decision-making to that encapsulated in the 'problem-solving model'.

The premise of this thesis is that these two models define the extremes of a continuum along which coverage decisions can be classified. In the analysis that follows, quantitative and qualitative methods will be deployed to assess where, along this continuum, different HTA bodies lie, and how this impacts on the HTA decision. To optimise and focus this research, a review of the literature was performed to identify the state of knowledge around the factors impacting on HTA funding decisions for pharmaceuticals.

## 2.2 Literature Review

In line with the theoretical models presented, the literature was examined to shed light on evidence, process, and context factors that impact on coverage decisions. The goal of the literature review was four-fold: (i) to critically examine the analyses performed of EU HTA bodies, (ii) to inform the selection of independent variables for analysis, the literature was examined to identify which evidence, process or context factors had been shown to impact on coverage decisions, (iii) to critically review the methods used to assess the relationship between evidence, process or context factors and coverage

decisions ; and (iv) to set the scene for the development of the analytical framework that will shape the methodology and analyses of this research.

### *2.2.1 Impact of evidence on HTA decision-making*

In the literature, both quantitative and qualitative analyses have been performed to understand how evidence is used in decisions to fund new pharmaceuticals. The literature agrees in its findings that evidence is usually a key component of the decision-making process, but that its impact is dependent upon the generalisability and quality of the evidence. Qualitative studies have looked at the issue of generalisability and quality of the evidence for decision-making, and how this can impact on the use of evidence in decision-making. In addition, qualitative and quantitative studies have looked at the relative impact of evidence on decision-making, in the face of competing factors and objectives.

The literature suggests that the degree of usefulness or accessibility of the data for decision-making is influenced by several factors. Accessibility of the data refers to the degree to which it is generalisable and of high quality and therefore useful to the decision-maker (Williams and Bryan 2007). The accessibility of the evidence is a necessary pre-requisite for evidence to have an impact on coverage decisions. Moreover, the availability of high quality data has been identified as a key factor for facilitating the use of evidence in the decision making process. Of similar importance is the need for HTA agencies to have access to data which are generalisable to their population. Thus, whether HTA expectations are met in terms of the quality and relevance of the data package submitted will depend on the quality and generalisability of clinical trial results, cost-effectiveness results and other factors such as whether or not the comparison of clinical efficacy is made with regard to the current standard of care. A recent development is the creation of an HTA adaptation toolkit and published in the literature, that assists in identifying the degree of relevance, generalisability, and quality of the submitted material to facilitate adaptation of HTA assessments between bodies (Turner et al. 2009; Chase et al. 2009).

One key characteristic of accessible evidence for decision-making is that it is relevant. Studies suggest that an evidence package which does not provide information that is relevant to HTA agencies will be less likely to obtain a positive funding decision

compared to a data package that meets the decision-maker's expectations. For example, in the USA, the FDA registrational file supporting gabapentin (for depression) was perceived to fulfil the quality expectations of the Federal Drug Administration (FDA) decision-makers and thus generated a positive recommendation from the FDA. However, when the evidence package supporting gabapentin was presented to state HTA decision-makers for inclusion in state formularies, a sub-set of regional HTA agencies did not accept the drug on their respective formulary list (Bloom, 2004). This divergence between the FDA and regional HTA decision-makers was, in part, attributed to the difference in objectives of the two decision-makers (one focused on public health objectives, the other on healthcare policy objectives), and differences in the data needed for them to make a decision. This example from the literature highlights the need for the evidence to be of relevance to its payer-audience, and suggests that the evidence that may be of relevance for regulatory/ registration purposes may not be relevant for HTA processes and vice versa.

Economic evaluation, and its accessibility, has been a recurrent theme in the literature. Several authors have discussed the barriers to the use of economic evaluations as part of HTA decision-making (Ross 1995; Williams and Bryan 2007; Bryan et al. 2007; Hoffman et al. 2000; Drummond et al. 1999; Bloom 2004). Lack of management of uncertainty around key parameters in the cost-effectiveness model was identified as a key factor that made evidence unhelpful to HTA agencies. Sculpher and Claxton (2005) highlight the need for managing the degree of uncertainty around key parameters of relevance for HTA decision-making – including the healthcare system objectives in which the HTA operates; how the product's incremental cost-effectiveness ratio (ICER) compares with the HTA agencies' threshold of willingness to pay; how the product compares to all relevant alternatives; and how and what costs are considered. McGuire et al. (2000) highlight the importance of understanding the objectives underlying economic evaluations, and in particular the objectives driving economic guideline development. In parallel, studies carried out in Europe and Australia identified similar barriers to the use of economic evaluation: the perception that the data was too complex; lack of generalisability; bias in the data; and untimely availability of data (Hoffman et al. 2000; Ross 1995; Bryan et al. 2007; Drummond, Schwartz et al. 2008). Thus, the literature suggests that the utility of economic evidence to the HTA decision-maker is influenced by several factors. It is further implied, but not directly assessed,

that lack of relevance and quality leads to non-use or reduced use of evidence from economic evaluations. Potentially, this could be extended to suggest that evidence which is not tailored to be useful to payer bodies will be less likely to influence the decision. Currently, the literature does not comment on barriers to accessibility for evidence other than economic evidence.

The use of cost-effectiveness evidence as part of HTA decision-making has been examined in several studies. Hoffman et al. (2000) conducted a European survey on the use of economic evaluations and found at the time that, on average, evidence from economic evaluations was considered by 30% of respondents. This is likely to have increased over time. Similar findings have been reported by Drummond et al. (1997) where use of economic evaluation was modest. More recently, in a qualitative study, Klarenbach et al. (2010) examined the role of economic evaluation in health technology assessments on anti-hypertensive therapy published by several HTA agencies, and noted that while economic evaluation was recognised as a factor in both HTA reports, the impact on the decision-making varied, as the reports differed in the conclusion of the technology that was regarded as 'most' cost-effective, and that this was driven principally by differences in the costs of the technology and its comparators.

Aside from cost-effectiveness considerations, the clinical and epidemiological evidence and the characteristics of the disease area for which it is indicated are important determinants of coverage decisions to grant or deny access. Greenhalgh et al. (2004) argues that the attributes of the innovation itself have a clear impact on its probability of diffusion – attributes include the relative advantage of the innovation vs. other existing products, compatibility, low complexity, and the nature of the knowledge required to implement the innovation. In countries where drug listings and formularies exist, a key criterion for inclusion in the list/formulary is the level of therapeutic value of the drug under consideration – lists are a means of excluding drugs perceived to have a low therapeutic value (Le Pen 1996; Fattore and Jommi 1998). The use of the ASMR rating system in France is an example of a system which bases the degree of funding awarded on the incremental therapeutic value demonstrated relative to standard of care.

Linked to the demonstration of added value is whether surrogate outcomes or hard endpoints are used. The demonstration of effectiveness through modification of

surrogate outcomes (e.g. blood pressure) are considered to be less convincing than the demonstration of effectiveness through modification of the frequency of outcomes/events (e.g. hypertension). Velasco Garrido et al. (2009) examined to what degree HTA recommendations utilised surrogate endpoints in their assessments. When examining the methodological guidelines from various HTA agencies, there were differences in the suggested use and appropriateness of surrogate outcomes in the appraisal process. Subsequently, in a sample of 140 HTA reports, 62% had utilised surrogate endpoints to support decision-making, although only 3% of recommendations were based exclusively on surrogate endpoints (Velasco Garrido et al. 2009). However, the authors did not examine to what degree the use of surrogate or hard endpoints impacted on the coverage decision. Nevertheless, their analysis highlights that the clinical and epidemiological evidence and the characteristics of the disease area for which it is indicated are key parts of the evidence used in making coverage decisions.

In addition to the nature of the endpoints within clinical trials, the nature of the disease itself, and particularly its severity, has been highlighted as a potentially important factor considered in coverage decisions (Drummond and Mason 2009; Bredesen 2003; Carlsson 2004; Clement et al. 2009). Both Norway and Sweden consider disease severity as an explicit factor in their decision-making processes (Bredesen 2003; Carlsson 2004). In addition, other healthcare systems have set up specific committees dedicated to specific disease areas, such as cancer, which are considered to be severe diseases. The Joint Oncology Drug Review committee in Canada is such an example (Mason and Drummond 2009). Furthermore, NICE has recently appended to its methodology guidelines specific advice to the appraisal committee on considerations to be taken into account when appraising so called ‘end-of-life’ technologies that may be life-extending (NICE 2009). Thus, aside from the nature of the clinical trial evidence, additional considerations linked to the disease characteristics are formal factors that are taken into account in coverage decisions.

The role of the evidence, relative to the role of other factors, has been examined in the literature. It is clear that HTA agencies/ payers have to consider a number of different, and often conflicting, factors when making decisions. Both qualitative and quantitative analyses have been performed that assess the relative impact of evidence in relation to other factors that may impact on the decision (Ross 1995; OECD 2005; Klarenbach et

al. 2010). Qualitative analyses have assessed the key factors that HTA agencies perceive to have an impact on their decisions. For example, in Ross (1995), 34 decision-makers within the funding process were interviewed. When asked which factors were important in decision-making, 44% cited economic appraisal or efficiency concerns; and 38% stressed the availability of relevant information and expert advice as having an impact on their decision. Similarly, in a survey of decision-makers within HTA systems across several countries, the OECD (2005) concluded that the evidence of efficacy/effectiveness (86%) and quality/safety (91%) were considered to be very important for funding decisions. In both studies, however, there were at least six other factors (such as equity or political considerations) which were ranked as important by the decision-makers. Thus, the available qualitative literature suggests that HTA agencies have to consider a number of different, and potentially conflicting, factors when making decisions. Therefore, it is necessary to consider an approach that attempts to analyse a set of factors simultaneously, rather than examines factors individually, as non-adjustment for the presence of other factors may lead to misinterpretation of the role of specific factors in decision-making.

Such qualitative analyses highlight two aspects of HTA decision-making: firstly, that while evidence is considered an important influential factor in decision-making by HTA agencies, decision-making is not 100% driven by evidence; secondly, that other factors, including process-related factors, efficiency and disease characteristics are also of concern for HTA agencies, within their decision-making process. From a quantitative point of view, international studies have examined the relative impact of evidence, in particular clinical and cost-effectiveness evidence, on decision-making (George et al. 2001; Clement et al. 2009; Lexchin and Mintzes 2008; Mason and Drummond 2009; Devlin and Parkin 2004; Dakin et al. 2006; Tappenden et al. 2007). George et al. (2001) examined, through development of a league table of reimbursed drugs by the Australian Pharmaceutical Benefits Advisory Committee (PBAC), whether PBAC decisions are consistent with the principle of maximum economic efficiency. Through the examination of 355 submissions made between 1991-1996, the authors report a statistically significant difference between the cost per life-year gained for drugs recommended for listing compared to those not recommended, through which it was inferred that the PBAC applies the criteria of economic efficiency in its decision making, and that the economic evaluation had an impact on the decisions made,

although other factors also influenced the decision (e.g. therapeutic value/need for the compound). However, the authors did not examine the impact of economic evaluation by adjusting for other factors. Thus, the relative importance of economic evaluation relative to the effect of other factors was not tested in this particular analysis.

In Europe, both agency-specific and cross-agency comparisons have been performed in the literature. With regard to single agency analyses, the literature has mostly focused on NICE, and on untangling, among various factors, those which appear to impact most on NICE decisions (Devlin and Parkin 2004; Dakin et al. 2006; Tappenden et al 2007; Mason and Drummond 2009). Devlin and Parkin (2004) use a binary choice analysis to identify the determinants of NICE recommendations, and consider a selection of explanatory variables including the cost per life year or per Quality Adjusted Life Year (QALY) gained; uncertainty regarding cost effectiveness result; budget impact from the NHS's perspective; the disease burden; and the availability (or not) of alternative treatments. The authors show that cost effectiveness, coupled with uncertainty and the burden of disease, explain NICE decisions better than cost effectiveness alone. Similarly, Tappenden et al. (2007), explore what factors NICE takes into account when making decisions through a stated preference binary choice experiment. Factors evaluated included preferences for incremental cost effectiveness, the degree of uncertainty surrounding incremental costs and health outcomes, the age of beneficiaries, baseline health-related quality of life (HR-QOL) and the availability of alternative therapies. Results suggest that a negative recommendation was more likely if the product under evaluation was associated with increases in the Incremental Cost Effectiveness Ratio (ICER), higher economic uncertainty, and the availability of alternative therapies ($p < 0.01$), suggesting that coverage decisions take multiple types of evidence into account.

To better reflect the reality of NICE decisions, Dakin et al. (2006) and Mason and Drummond (2009) depart from a binary modelling approach. Instead of assuming that NICE operates with a binary choice system, both studies categorise NICE coverage decisions into three coverage levels: recommendation for use, recommendation for restricted use and no recommendation for use in the National Health Service (NHS). In Dakin et al.'s (2006) analysis, the NICE appraisal process was modelled as a single decision between the three categories. Multinomial logistic regression techniques were

used to evaluate the impact of: quantity/quality of clinical evidence; cost-effectiveness; existence of alternative treatments; budget impact and technology type. Results indicated that evidence was an important driver of NICE decisions: interventions supported by a greater number of randomised trials and those with more systematic reviews were less likely to be rejected. Furthermore, interventions with 'higher' cost-effectiveness ratios were more likely to be rejected rather than recommended for restricted use.

The literature comparing coverage decisions and factors driving variation in coverage decisions across agencies has generally included NICE and/or SMC, as well as the Canadian and Australian HTA agencies (Clement et al. 2009; Lexchin and Mintzes 2008; Barbieri et al. 2009). Two studies (Barbieri et al. 2009; Lexchin and Mintzes 2008), attempt to examine factors explaining coverage decision patterns within their sample, but neither applied methods to assess the contribution of the variables of interest while adjusting for potential confounding factors. Barbieri et al. (2009), in their comparison of NICE and the SMC coverage decisions, attempted to explain differences in coverage decisions through the fact that NICE uses third party assessment, while SMC does not. However, the lack of sample size impeded the authors from, firstly, demonstrating the presence of statistically significant differences in coverage decisions between the two bodies, and secondly, from demonstrating if these differences were driven by third party technology assessment processes, rather than other factors (e.g. use of clinical/economic evidence, agency's mission, other process elements e.g. inclusion of patient groups etc). The coverage decisions of the SMC were also compared with the Australian Pharmaceutical Benefits Advisory Committee (PBAC) and the Canadian Common Drug Review (CDR) in a study by Lexchin and Mintzes (2009). The authors found statistically significant differences in coverage decision patterns between the bodies; however, the analysis of the factors driving such differences was limited to qualitative assessment of a subset of individual technology appraisals to identify potential factors driving the observed discrepancy in coverage decisions. Possibly due to limited data and small sample size, Lexchin and Mintzes (2009) did not test in a quantitative and comprehensive fashion the degree to which different factors could be contributing to diverging patterns of coverage decisions between the agencies.

Of the available studies in the literature, Clement et al. (2009) provide the only analysis that aimed to explain differences in coverage decisions between NICE, PBAC and CDR through the systematic collection of data on a range of factors, and their analysis through quantitative techniques. For each appraisal, data was collected on a range of clinical, economic and process factors. Clinical variables included the level of clinical uncertainty, weight of the clinical evidence, the study endpoint of the clinical trial, the goal of the treatment (quality of life enhancement, life extension, or both). Economic criteria that were considered were the type of cost-effectiveness evidence provided and the level of economic uncertainty. Through descriptive statistics, each of the variables was compared across the three agencies. This analysis suggested that there were statistically significant differences between the agencies in the types of technologies appraised (a higher percentage of technologies appraised by CDR were aimed at improving both quality of life and life expectancy compared to the other agencies), while NICE appeared to appraise technologies with a higher proportion of clinical end points in their clinical trials (as opposed to surrogate endpoints). There were also significant differences in the type of economic evidence considered by the three agencies – with NICE most likely to consider cost utility analyses, and PBAC and CDR more likely than NICE to consider cost minimization analyses ($p<.001$). There did not appear to be statistically significant differences in the level of economic uncertainty of the technologies appraised by the three agencies.

The results of the Clement et al. (2009) analysis suggest that there are statistically significant differences in the nature of the coverage decisions made by NICE, PBAC and CDR. And additionally, the results show that there are differences in the characteristics of the technologies appraised by the agencies, and the evidence that supports them. However, the study is not able to clarify the relative importance of each of the examined factors relative to the other, as a regression model was not attempted to allow for the adjustment of the effect of individual factors towards one another. Nor was the study able to provide direct evidence of the impact of variation in a specific variable leading to variation in coverage decisions.

In summary, the literature on the impact of evidence on coverage decisions described above focuses on the degree to which evidence related to the medicine under review (whether clinical, economic or otherwise) can impact on HTA coverage decisions. The

literature suggests that, while evidence is an important component that influences the decision outcome, it is necessary to look beyond the evidence base and to consider other non-evidence based factors that impact on decisions – in particular, the processes that guide decision-making, as well as the socio-economic-political context in which these decisions are made.

### *2.2.2 Process-related factors that impact on the payor decision*

> ". . . it doesn't matter what your decision is when you come to the end. It's the process that's got to have been absolutely rigorous." (Pharmaceutical advisor, interview 4 in Wirtz et al. 2005 p. 335)

The literature examining the HTA appraisal process is primarily qualitative in nature. It provides insights into a number of process-related factors that can potentially influence coverage decisions including: (i) the approach to decision-making and criteria considered as part of the decision-making process; (ii) composition of the decision-making committee; and (iii) key components that make up the decision-making process.

It can be hypothesised that the approach to HTA decision-making, or, in other words, the objective that the appraisal process strives for, shapes the HTA process and subsequent decisions. In practice, most decision-making processes are attempting to address numerous objectives simultaneously, with HTA being an approach that endeavours to consider several different components in one assessment, including costs and effectiveness but also clinical data, epidemiological data and disease characteristics, equity, social considerations, patient perspective, industrial policy objectives and so forth. Evidence of the impact of different HTA approaches/ objectives on coverage decisions is limited. Al et al. (2004) aimed to assess the underlying objectives that HTA agencies were striving for when making their decisions. This was done through semi-structured interviews with four stakeholders within the Dutch decision-making system. Several HTA objectives were tested: to maximize health gain for a fixed budget (efficiency); to maximize net health benefit with and without weighted health effects; decreasing marginal value with/without weighted health effects; equity considerations by age or socio-economic status . These seven goals were presented to the interviewees within seven potential budgetary constraints. Interviewees were asked to rank the goals

in terms of which they would prioritise first, and across the four interviewees there was wide variation.

While the sample size in Al et al. (2004) is too small to allow generalisability, it suggests that HTA agencies do have formalized objectives through which they make their decisions and that therefore, understanding the objectives may help to better understand coverage decisions. For example, George et al. (2001) examined whether PBAC decisions are consistent with the principle of maximum economic efficiency, and report a statistically significant difference between the cost per life-year gained for drugs recommended versus those not recommended, through which it is inferred that the PBAC applies the criteria of economic efficiency in their decision making. Menon et al. (2005) in Canada, evaluated spatial variation in the availability of cancer drugs across ten Canadian provinces, and assessed inter-provincial variations in the processes that govern those decisions. The authors reported that only 7 of 115 cancer drugs evaluated were available in all ten provinces, and indirectly link this variation in access to the decision-making process adopted, and as a result, the variation in the relative importance of different components in their decision-making.

How and by whom the objective or approach towards decision-making is implemented can be supposed, at least theoretically, to have an impact on decision outcomes. Indeed, studies suggest that the composition of the decision-making committee can impact on the HTA decision. HTA decision-making bodies tend to be composed of a mix of individuals from different roles and backgrounds. In a review of decision-making processes Vuorenkoski et al. (2008) concluded that experts such as pharmacists or physicians were key stakeholder groups with considerable influence on the HTA decision-making process. Menon et al. (2005) analysed the variation in availability of oncology therapies across provinces in Canada and found that there were significant differences between HTA bodies in the frequency of meetings, composition of the committee in terms of absolute size and characteristics of the members (i.e. clincians, staff of HTA agencies, pharmacists, health economists etc). Similarly, the OECD (2005) surveyed the types of participants involved in decision-making across several countries: in general, government officials (67%), health-care managers (76%) and academics/technical experts (67%) were most commonly involved in HTA decision-making. Patients/consumer groups, politicians and industry representatives were the

least involved (<25%). Both studies show that variation exists in the composition of HTA decision-making bodies and their operational characteristics. However, neither study attempted to directly link the differences in the composition of the decision-bodies used by different HTA agencies to differences in coverage decisions across provinces/HTA bodies.

Iversen and Vondeling (2007), on the other hand, do consider the impact of actors within the reimbursement process in Denmark on HTA decision making and diffusion of technologies. In a case-study on the reimbursement of glitazones in Denmark, Iversen and Vondeling (2007) provide a detailed analysis of the different actors involved in the reimbursement process including the ministry of health, the Danish Medicines Agency, the counties and non-governmental organizations, including the industry and patient groups. They assess, for this particular case study, how the reimbursement decision came about as a result of the interaction between the actors identified in the process. The analysis highlights the importance of considering the extent of involvement of various stakeholders and how this impacts on the HTA decision. However, as the analysis was based on a single case study in Denmark, its generalisability is limited.

The components of the appraisal process, as distinct from the participants within the process, have been examined in the literature and their potential impact on coverage decisions has been highlighted (Dakin et al. 2006; Barbieri et al. 2009; Vuorenkoski et al. 2008). The components of the appraisal process refer to the different elements which are evaluated as part of the appraisal process. Thus, for instance, whether economic evaluation is included, and whether patient submissions are allowed are two components which may or may not be part of an appraisal process.

Research to assess how variation in the presence or absence of these components can impact on coverage decisions is largely limited to qualitative analyses that describe variation in use of specific components within HTA appraisals, but are not able to demonstrate a link between such variation and differences in coverage decisions.
In a qualitative review, severity of disease and past decisions made by the agency appeared to be important components of decision-making (Vuorenkoski et al. 2008), although evidence of direct impact on coverage decision was lacking. Barbieri et al.

(2009) assessed to what degree the use of third party technology assessment (by external academic groups) could impact on coverage decisions. This involved comparison of NICE and the SMC – in the former, third party evidence is used, while this is not the case in the SMC appraisal process. Case studies and examples where third party assessment appeared to be important in explaining differences in recommendations between NICE and SMC were highlighted. However, sample size limitations meant that the impact of third party assessment on NICE and SMC coverage decisions could not be examined more formally.

Additional research across other European and non-European HTA bodies confirm that the elements included as part of the HTA process vary by country and by type of intervention. The OECD (2005) provides an overview of the different types of evidence considered in Health Technology Assessments for 5 different types of interventions (PET, Hepatitis C genotyping, Telemedicine, Prostrate cancer screening and stroke technologies). While none of these interventions is a pharmaceutical, the results of this study give a flavour of the types of evidence considered by HTA agencies in these HTAs. Evidence considered includes information on efficacy/effectiveness, quality/safety, cost-effectiveness, and burden of disease in population. In addition to the evidence base varying by type of intervention, it also appears to vary by country (OECD 2005). For example, in the HTA of PET, a survey of HTA agencies across eight countries revealed that evidence on effectiveness and safety was consistently used across countries, while evidence of cost-effectiveness and total cost burden was considered by approximately half of the countries. No two countries were alike in the type of evidence considered. Similarly, Draborg et al. (2005) reviewed 11 HTA bodies in 2002, and identified differences in the scope and nature of the country-specific HTA processes. The clinical component of the HTA assessment was most consistently considered, while economic, patient and organisation-level impact was less consistently analysed. In about half of the HTA bodies evaluated, HTA processes included an economic evaluation, often in the form of a cost-effectiveness analysis. Since 2002, more HTA bodies and pricing and reimbursement (P&R) processes have adopted economic evaluations of new pharmaceutical compounds. Neither study assessed to what extent differences in the data used between countries could explain differences in coverage decision.

In an assessment of NICE guidance, interventions supported by patient group submissions increased the probability of a recommendation for routine rather than restricted use (Dakin et al. 2006). This represents the only identified analysis in which the absence or presence of a component of the appraisal process was linked to coverage decisions. It was not, however, able to comment on to what degree variability in the use of this component could help explain variability in coverage decisions between HTA agencies.

Related to the review of the characteristics of HTA assessments across HTA bodies in Europe, a body of literature is available that focuses specifically on the analysis of the extent to which economic evaluations are part of the HTA decision-making process. Bloom (2004) examined the use of formal benefit/cost evaluations by health care decision makers (n=104) in the USA, UK, France and Sweden. 42.1% of respondents reported using formal economic analyses in their funding decisions. Kanavos et al. (2000) consider the place of cost-effectiveness evaluation in the decision-making process, and document the variations observed in Portugal, Netherlands, Finland and the UK in how the cost-effectiveness component is applied within the appraisal process.

In addition to assessing the presence or absence of economic evaluations within HTA, studies have examined the components within the economic component – i.e. the methodologies used for economic evaluations. Methodologies for economic evaluation have different components within them, and the incremental cost-effectiveness ratio (ICER) produced by an economic evaluation is dependent in part on the outcomes included and in part on the methodology with which it is derived – i.e. perspective, time horizon, discount rate, costs to be included, choice of comparator, patient population etc (Kanavos et al. 2000; Hjelmgren et al. 2001; Drummond 2003; Sculpher and Drummond 2006). Hjelmgren et al. (2001) show that, across countries, the level of methodological agreement between Health Economic guidelines ranged from 40-100%. Sculpher and Drummond (2006) examine variation in 27 guidelines for economic evaluation across countries, highlighting considerable variation in the methods recommended on choice of comparator, approach sensitivity analysis, identification of relevant evidence, etc. Barbieri et al. (2005) examined 46 inter-country drug comparisons of cost-effectiveness studies, to assess the variability of the cost-effectiveness results and note that the variability observed was not systematic, thus

highlighting the importance of the country-specific cost-effectiveness threshold as a key determinant of the outcome of the decision. Despite identifying variability in the results of cost-effectiveness analyses for the same technology, the authors do not explore how this may explain variability in HTA recommendations. It is of interest to examine further how differences in the use and methodology of economic evaluation may be partly accountable for differences in HTA recommendations between countries/assessment bodies.

### *2.2.3 Healthcare and welfare context – impact on HTA coverage decision*

The literature has shown that coverage decisions are influenced by macro factors, such as healthcare spending per capita, societal willingness to pay, the structure of the healthcare system, as well as ethical and social considerations (Packer et al. 2006; Owens 1998; Buxton 2005; Bryan et al. 2007; Ross 1995). This section focuses on the *acceptability* of evidence within the context of the decision-making process (Williams and Bryan 2007). Thus, the notion of acceptability from a macro point of view can be analysed by considering two distinct types of acceptability: structural/ institutional acceptability and ethical/political acceptability (Williams and Bryan 2007).

Packer et al. (2006) consider the impact of specific factors on the differential diffusion of six innovations in ten countries. Results of the analysis showed that there were some systematic country trends: i.e. a tendency for a selection of countries to have a high up-take, including Sweden, Switzerland, Canada and Norway, and other countries to have a tendency for low up-take, including UK, Spain and Denmark. Of the variables considered, macro economic factors measured in terms of health spending per capita appeared to be associated with an increased diffusion of the health technology, while an above average funding from taxation appeared to be linked to a reduced rate of diffusion of two health technologies (verteporfin and interferon beta). While the analysis did not focus on the impact of socio-economic factors on HTA coverage decisions, the analysis of different patterns of diffusion suggests that such factors could be of relevance when attempting to explain patterns of coverage decisions.

Differences in societal willingness to pay for a healthcare benefit (such as an additional QALY) will impact on the cost-effectiveness threshold used in decision-making and thus may also impact on coverage decisions. There is considerable literature illustrating

the variation of the cost-effectiveness threshold across time and geography, and the aim here is not to provide a comprehensive overview of this set of literature, but to highlight that the cost-effectiveness threshold does vary between HTA bodies (both within and between countries and across time) and the existence of such a threshold can have an impact on the final HTA decision to recommend or reject a pharmaceutical for public funding. Owens (1998) considers factors that influence the interpretability of cost-effectiveness analyses, and raises the question of how to determine the cut-off point at which an ICER should be considered cost-effective or not. Buxton (2005), in his discussion of the 'implicit' cost utility thresholds in the first 39 NICE appraisals, discusses the principles that may be used to determine the appropriate and explicit cost-effectiveness threshold. The GDP of a country is considered as a potential method of deriving a societal value for an added quality-adjusted-life-year, within the economic constraints of that society. Ultimately, differences in the use of cost-effectiveness thresholds are expected to impact on the outcome of HTA coverage decisions, although currently the literature available has not explored this particular dimension. Therefore, it is of interest to consider differences between European Member States in how the cost-effectiveness threshold is determined and applied to coverage decisions about the public funding of pharmaceuticals.

Several authors have discussed the acceptability barriers, particularly with reference to the use of economic evaluations as part of HTA decision-making (Ross 1995; Bloom 2004; Williams and Bryan 2007; Bryan et al. 2007; Hoffman et al. 2000; Drummond et al. 1999; Kanavos et al. 2000; Sculpher and Claxton 2005). One of the barriers highlighted in the literature is the perceived difficulty by HTA agencies in converting a theoretical economic benefit into reality within their healthcare system, and secondly, scepticism about whether projected savings could actually be realized (Ross 1995; Hoffman et al. 2000).

The HTA decision and the extent to which it is purely evidence-based is also influenced by how it marries with other key considerations in decision-making, such as political, social or ethical considerations (Bryan et al. 2007; Ross 1995). For example, in a qualitative study examining the factors impacting on the use of cost-effectiveness analyses in the NICE decision-making process, the difficulty of reconciling CEA analysis with equity considerations, such as disease severity, was raised (Bryan et al.

2007). Diverging views were expressed from NICE committee members with different backgrounds and expertise. Some criticized CEA from a methodological standpoint, pointing out their belief that currently no methodological approach was available to take equity into consideration within CEA analyses. Others highlighted the inconsistent approach with which equity principles are applied. Situations were described when disease severity may carry more weight in the decision-making process and override or neutralize the evidence:

> "The fact that it is an important disease that causes death focuses the mind a little more than perhaps some other technologies we've looked at where there may be good randomised clinical trials but sometimes it's difficult to judge the relative merits of the technology" (Bryan et al. 2007 p. 189).

Thus, disease severity/equity considerations can influence the HTA decision by impacting on the evidence used in the process, and impact on the decision itself.

### 2.3 Gaps in the literature and implications for this research

The literature was examined and analysed to (i) to critically examine the analyses performed of EU HTA bodies, (ii) to inform the selection of independent variables for analysis, the literature was examined to identify which evidence, process or context factors had been shown to impact on coverage decisions, (iii) to critically review the methods used to assess the relationship between evidence, process or context factors and coverage decisions ; and (iv) to set the scene for the development of the analytical framework that will shape the methodology and analyses of this research. In summary, the review of the literature suggests there are limitations in terms of their scope of EU HTA bodies analysed, analytical methodology adopted to examine HTA coverage decisions and comprehensiveness of the variables considered in analyses of HTA coverage decisions. These limitations and the implications for this research are examined further below.

A key limitation identified in the literature is the focus on NICE as an example of HTA in Europe, despite the fact that in recent years there has been a growth in the use of HTA across EU Member States (OECD, 2005). The scope of the available literature has either focused on agency-specific analyses (primarily NICE-related), and any cross agency analyses have included Anglophone agencies, namely NICE and SMC, alongside CDR in Canada and PBAC in Australia. Multiple European HTA agencies have been compared and contrasted in literature examining the characteristics and

differences in processes between HTA bodies, but without linking such differences in characteristics and processes to explaining variation in coverage decisions (Sorenson 2008; Hutton et al. 2006). For example, a comparative assessment of the characteristics of different HTA agencies was provided and proposed by Hutton et al. (2006). While the framework proposed represents a useful starting point for developing an analytical framework and for identifying variables of relevance for a comparative analysis, it is not designed to analyse to what extent the use of HTA (i.e. the characteristics of the technology under evaluation and the process by which it is evaluated) impacts on the HTA decision to grant, restrict or deny public funding for new technologies. In response to this current gap in the literature, this thesis aims to contribute by broadening the scope of the available literature by analyzing the factors influencing HTA decision-making in a selection of European HTA agencies.

A second limitation observed in the literature on coverage decisions by HTA agencies is the consideration of a limited range of variables/factors. The literature examining factors driving decision-making has to a large extent focused on specific types of factors (e.g. papers focusing on economic factors, others on process factors etc.). Few have combined factors (e.g. Clement et al. 2009), and there are no studies which have combined clinical, economic, process, disease and socio-economic context variables into a single analysis, extracted directly from HTA reports. While process and context-related factors have been identified as potentially important factors influencing coverage decisions, and differences in these factors between HTA bodies have been described, few authors have in fact made a link between differences in process and context factors, and coverage decisions. Several surveys and qualitative analyses have shown that HTA agencies consider a wide range of evidence and that there is variation in the components of the decision-making process. However, little research has been performed to assess how variation in the presence or absence of these components can impact on coverage decisions. With regard to the impact of process-related factors, such as the composition of decision-making committees, while the literature shows that variation exists in the composition of HTA decision-making bodies and their operational characteristics, there is no attempt to directly link the differences in composition used by different HTA bodies to differences in coverage decisions across provinces/HTA bodies. This thesis aims to contribute by considering a broad range of clinical, economic, process and context variables in the analysis of factors driving coverage decisions.

Another important gap observed in the literature is related to the analytical methodology adopted to assess coverage decisions and relevant factors. In general, the analyses adopted have been primarily qualitative or have adopted descriptive quantitative methodologies. The study comparing NICE, PBAC and CDR was descriptive in nature (Clement et al. 2009). Such descriptive analytical techniques make it difficult to interpret the relative contribution of each factor, given the absence of adjustment for other factors in the analysis. Models of NICE decision-making have been developed for NICE, but not for other European HTA agencies, neither as single agency analyses nor as comparative analyses across several European HTA bodies. This thesis aims to address this gap by creating a bespoke dataset of HTA coverage decisions from four different HTA bodies in Europe over a five-year period and utilising statistical methods of analysis to assess the relative contribution of a comprehensive range of factors on coverage decisions both within each HTA body and across HTA bodies.

Based on the literature review presented, and on the identified gaps in the literature, the section that follows will aim to present the analytical framework that will shape the methodology, structure and analyses performed as part of this thesis.

## 2.4 Analytical Framework

The literature suggests that while clinical and economic evidence is an important component that can influence HTA coverage decisions, an analytical framework would need to look beyond the evidence base and also consider other non-evidence based factors that can impact decisions – in particular the processes that guide decision-making, as well as the socio-economic-political context in which these decisions are made. An analytical approach has been developed to assess and compare HTA decision-making across healthcare systems, and to identify those factors that can explain HTA decision-making patterns observed within and between HTA bodies across EU Member States. The principles upon which the framework is based are derived from the theoretical concepts and empirical evidence identified from the review of the literature presented earlier.

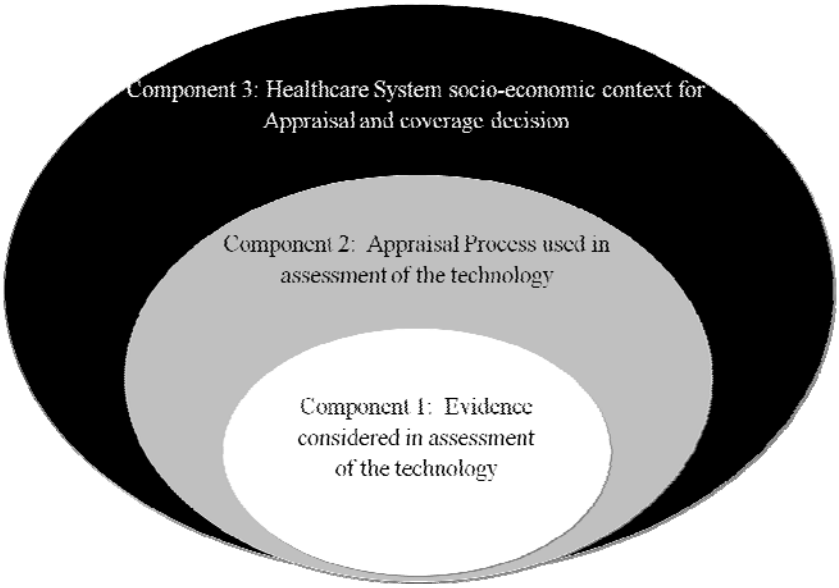
From a theoretical perspective, the movement of evidence-based medicine in the realm of policy making has meant that evidence-based policy-making has become an

important movement. Payers, like other policy decision-makers have moved towards incorporating the elements of evidence-based policy making into their decision-making. Weiss (1979) introduced two models that can help understand how evidence is incorporated in decision-making: the 'problem-solving model' and the 'interactive' model, the former being an evidence-driven model while the latter is driven by interactions between different stakeholders and factors within the decision-making process, of which evidence is only one component. When thinking about the transferability of these theoretical models to the analysis of HTA decision-making in real-life healthcare systems, it is necessary to analyse the extent to which evidence is used in HTA decision-making, and to what extent process and context factors influence coverage decisions.

In alignment with the theoretical concepts identified above, three "streams" of research become apparent from the literature. Firstly, research on the impact of evidence on the HTA decision. Secondly, research on the decision-making process itself, rather than on the technology (e.g. whether economic evaluation is a component of the decision-making process or not). Thirdly, reference in the literature is made to the impact of the overall healthcare and welfare characteristics on HTA decision-making (e.g. the impact of healthcare expenditure levels and the health policy priorities being reinforced by the Ministry of Health at the time of HTA decision-making). Given these themes observed in the literature, the analytical framework is based on a three-pronged approach and employs three components of analysis (Fig 2.2):

- Component 1: Technology evidence-base characteristics: assessment of the technology's supporting evidence
- Component 2: Process characteristics – examines the process with which technologies are evaluated
- Component 3: Socio-economic characteristics, which examines the socio-economic and political context for decision-making

**igure 2.2 Three-component framework for evaluating outcomes of the HTA appraisal process**




Component 1 focuses on analyzing the availability and quality of the data inputs related to the technology and the disease area that go into specific health technology appraisals. The second component focuses on capturing the characteristics of the decision-making process itself, not specific to any one technology assessment. The third component takes a macro approach and aims to assess overall systemic factors, including social, economic and political factors that could influence HTA decision making.

Ultimately, the approach employed in this thesis aims to examine empirically to what extent different factors play a role in HTA decision making, and where coverage decisions lie within the spectrum moving from a tailored, systematic evidence-driven approach to a multi-factorial approach to decision-making on coverage. The three-component analytical framework will be used to structure thinking around the appropriate methodology to address the research question, in particular regarding the research methods, variables for consideration, data sources and consideration of limitations.

## 2.5 References

Al, M. J., T. Feenstra, and W. B. Brouwer. 2004. Decision makers' views on health care objectives and budget constraints: results from a pilot study. *Health Policy* 70 (1):33-48.

Anell, A, and U Persson. 2005. Reimbursement and clinical guidance for pharmaceuticals in Sweden. Do health-economic evaluations support decision making? *Eur J Health Econom* 50:274-279.

Barbieri, M, MF Drummond, R Willke, J Chancellor, B Jolain, and A Towse. 2005. Variability of Cost-Effectiveness Estimates for Pharmaceuticals in Western Europe: Lessons for Inferring Generalizability. *Value in Health* 8 (1):10-23.

Barbieri, M., N. Hawkins, and M. Sculpher. 2009. Who does the numbers? The role of third-party technology assessment to inform health systems' decision-making about the funding of health technologies. *Value Health* 12 (2):193-201.

Bloom, BS. 2004. Use of Formal Benefit/Cost Evaluations in Health System Decision Making. *The American Journal of Managed Care* 10 (5):329-335.

Bredesen, G. 2003. Organization of health care systems and financing of pharmaceuticals in Norway, including cost containments. *J Pharmaceut Finance Econ Pol* 12:351–60.

Brousselle, A., and C. Lessard. 2011. Economic evaluation to inform health care decision-making: Promise, pitfalls and a proposal for an alternative path. *Social Science & Medicine* 72(6): 832-839.

Bryan, S., I. Williams, and S. McIver. 2007. Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. *Health Econ* 16 (2):179-93.

Buxton, M. 2005. How much are health-care systems prepared to pay to produce a QALY? *Eur J Health Econom* 6, 285–287.

Carlsson, P. 2004. Health technology assessment and priority setting for health policy in Sweden. *Int J Technol Assess Health Care* 20 (1):44-54.

Chase, D., Rosten, C., Turner, S., Hicks, N., and Milne, R., 2009. Development of a toolkit and glossary to aid in the adaptation of health technology assessment (HTA) reports for use in different contexts. Health Technology Assessment 13 (59):1-118

Clement, F. M., A. Harris, J. J. Li, K. Yong, K. M. Lee, and B. J. Manns. 2009. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 302 (13):1437-43.

Commission de la Transparence. 2007. SPRYCEL 20 mg, comprimé pelliculé, plaquette thermoformée (377 637-9). CT 4070. Avis 14 Mars 2007. http://www.has-sante.fr/portail/upload/docs/application/pdf/ct-4070_sprycel_.pdf

Dakin, HA , NJ Devlin, and IAO Odeyemi. 2006. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. *Health Policy* 77:352-367.

Daniels, N., and Sabin, JE. 1997. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philosophy and Public Affairs* 26:303-350

Devlin, N., and D. Parkin. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 13 (5):437-52.

Draborg, E, BP Poulsen, and M Horder. 2005. International Comparison of the definition and the practical application of health technology assessment. *International Journal of Technology Assessment in Health Care* 21 (1):89-95.

Drummond, M., Cooke, J., and Walley, T. 1997. Economic evaluation under managed competition: evidence from the UK. *Social Science and Medicine* 45(4):583–95.

Drummond,. M, Dubois, D., and Garattini, L., Horisberger, B., Jonsson, B., Kristianses, I.S., Le Pen, C., Pinto, C.G., Poulsen, P.B.O., Rovira, J. Rutten, F., von der Schulenburg, M., and Sintonen, H. 1999. Current trends in the use of pharmacoeconomics and outcomes research in Europe. *Value Health* 2:323-332.

Drummond, MF. 2003. The use of health economic information by reimbursement authorities. *Rheumatology* 42 (Suppl. 3):iii60-iii63.

Drummond, M., R. Brown, A.M. Fendrick, P. Fullerton, P. Neumann and R. Taylor *et al.*. 2003. Use of pharmacoeconomics information: report of the ISPOR Task Force on use of pharmacoeconomic/health economic information in health-care decision making, *Value in Health* 6 (4):407–416.

Drummond, M., and A. Mason. 2009. Rationing new medicines in the UK. *BMJ* 338:a3182.

Drummond, M. F., J. S. Schwartz, B. Jonsson, B. R. Luce, P. J. Neumann, U. Siebert, and S. D. Sullivan. 2008. Key principles for the improved conduct of health

technology assessments for resource allocation decisions. *Int J Technol Assess Health Care* 24 (3):244-58; discussion 362-8.

Fattore, and Jommi. 1998. The new pharmaceutical policy in Italy. *Health Policy* 46:21-41.

George, B, A Harris, and A Mitchell. 2001. Cost-Effectiveness Analysis and the Consistency of Decision Making. *Pharmacoeconomics* 19 (11):1103-1109.

Greenhalgh, T, G Robert, F Macfarlane, P Bate, and O Kyriakidou. 2004. Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *The Milbank Quarterly* 82 (4):581-629.

Hasle-Pham, E., B. Arnould, H. M. Spath, A. Follet, G. Duru, and P. Marquis. 2005. Role of clinical, patient-reported outcome and medico-economic studies in the public hospital drug formulary decision-making process: results of a European survey. *Health Policy* 71 (2):205-12.

Hjelmgren, J, F Berggren, and F Andersson. 2001. Health Economic Guidelines - Similarities, Differences and Some Implications. *Value in Health* 4 (3):225-250.

Hoffmann, C., and J. M. Graf von der Schulenburg. 2000. The influence of economic evaluation studies on decision making. A European survey. The EUROMET group. *Health Policy* 52 (3):179-92.

Hutton, J, C McGrath, JM Frybourg, M Tremblay, E Bramley-Harker, and C Henshall. 2006. Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal of Technology Assessment in Health Care* 22 (1):10-18.

Iversen, PB, and H Vondeling. 2006. Reimbursement Decision-Making and Prescription Patterns of Glitazones in Treatment of Type 2 Diabetes Mellitus Patients in Denmark. *Health Care Anal* 14:79-89.

Lexchin, J., and B. Mintzes. 2008. Medicine reimbursement recommendations in Canada, Australia, and Scotland. *Am J Manag Care* 14 (9):581-8.

Kanavos, P, P Trueman, and A Bosilevac. 2000. Can economic evaluation guidelines improve efficiency in resource allocation? The cases of Portugal, the Netherlands, Finland, and the United Kingdom. *International Journal of Technology Assessment in Health Care* 16 (4):1179-1192.

Klarenbach, S.W., McAlister, F.A., Johansen, H., Tu, K., Hazel, M., Walker, R.,

Zarnke K.B., and Campbell, N.R.C. 2010. Identification of factors driving differences in cost effectiveness of first-line pharmacological therapy for uncomplicated hypertension. *Canadian Journal of Cardiology* 26 (5):e158-e163.

Le Pen, C. 1996. Drug pricing and reimbursement in France. Towards a new model? *Pharmacoeconomics* 10, 26-36.

Mason, A., and Drummond, M. 2009. Public funding of new cancer drugs: Is NICE getting nastier? *European Journal of Cancer* 45:1188-1192.

McGuire, A., S. Morris, and M. Raikou. 2000. Where are the economic guidelines coming from? *International Journal of Health Technology Assessment*,16:976-986.

Menon, D, T Stafinski, and G Stuart. 2005. Access to drugs for cancer: Does where you live matter? *Can J Public Health* 96 (6):454-8.

Mossialos, E., Mrazek, M., and Walley, T. 2004. *Regulating pharmaceuticals in Europe: striving for efficiency, equity and quality*. Maidenhead, UK: Open University Press.

National Institute for Health and Clinical Excellence (NICE). 2008. *NICE technology appraisal guidance 143 - Adalimumab, etanercept and infliximab for ankylosing spondylitis*. London: National Institute for Health and Clinical Excellence, 2008.

———. 2009. Supplementary advice to the Appraisal Committees. http://www.nice.org.uk/media/E4A/79/SupplementaryAdviceTACEoL.pdf. Viewed on 11 January 2011.

OECD. 2005. *Health Technologies and Decision Making*. The OECD Health Project, OECD Publishing.

Owens, D. 1998. Interpretation of cost-effectiveness analyses. *JGIM* 13, 716-717.

Packer, C., S. Simpson, and A. Stevens. 2006. International diffusion of new health technologies: a ten-country analysis of six health technologies. *Int J Technol Assess Health Care* 22 (4):419-28.

Permanand, G. 2006. *EU pharmaceutical regulation : the politics of policy-making / Govin Permanand*. Manchester: Manchester University Press

Ross, J. 1995. The use of economic evaluation in health care: Australian decision makers' perceptions. *Health Policy* 31 (2):103-10.

Sculpher, M, and K Claxton. 2005. Establishing the Cost-Effectiveness of New Pharmaceuticals under Conditions of Uncertainty - When is There Sufficient Evidence? *Value in Health* 8 (4):433-446.

Sculpher, M, and MF Drummond. 2006. Analysis Sans Frontieres - Can We Ever Make Economic Evaluations Generalisable Across Jurisdictions? *Pharmacoeconomics* 24 (11):1087-1099.

Scottish Medicines Consortium (SMC). 2007. dasatinib, 20mg, 50 mg, 70 mg tablets (Sprycel®) No. (370/07) 6 April 2007.

Sibbald, S. L., P. A. Singer, R. Upshur, and D. K. Martin. 2009. Priority setting: what constitutes success? A conceptual framework for successful priority setting. *BMC Health Serv Res* 9:43.

Sinclair, S., N. A. Hagen, C. Chambers, B. Manns, A. Simon, and G. P. Browman. 2008. Accounting for reasonableness: Exploring the personal internal framework affecting decisions about cancer drug funding. *Health Policy* 86 (2-3):381-90.

Sorenson, C., Drummond, M., and Kanavos, P. 2008. Ensuring value for Money in Health Care: the role of HTA in the European Union. Cornwall: World Health Organization 2008, on behalf of the European Observatory on Health Systems and Policies.

Tappenden, P., J. Brazier, J. Ratcliffe, and J. Chilcott. 2007. A stated preference binary choice experiment to explore NICE decision making. *Pharmacoeconomics* 25 (8):685-93.

Turner, S., D. L. Chase, R. Milne, A. Cook, N.J. Hicks, C. Rosten, L. Payne, S. Coles and E. Bell. 2009. The health technology assessment adaptation toolkit: Description and use. *International Journal of Technology Assessment in Health Care* 25: 37-41.

Velasco Garrido, M., and S. Mangiapane . 2009. Surrogate outcomes in health technology assessment: An international comparison. *International Journal of Technology Assessment in Health Care* 25:315-322

Vogel, R. J. 2004. Pharmaceutical pricing, price controls, and their effects on pharmaceutical sales and research and development expenditures in the European Union. *Clin Ther* 26 (8):1327-40; discussion 1326.

Vuorenkoski, L., H. Toiviainen, and E. Hemminki. 2008. Decision-making in priority setting for medicines--a review of empirical studies. *Health Policy* 86 (1):1-9.

Weiss, C. 1979. The Many Meanings of Research Utilization. *Public Administration review* 39 (5):426-431.

Williams, I., and S. Bryan. 2007. Understanding the limited impact of economic evaluation in health care resource allocation: a conceptual framework. *Health Policy* 80 (1):135-43.

Wirtz, V., A. Cribb, and N. Barber. 2005. Reimbursement decisions in health policy--extending our understanding of the elements of decision-making. *Health Policy* 73 (3):330-8.

# 3 Methods

The objective of this thesis is to assess the factors that drive HTA coverage decisions to recommend, restrict or not recommend technologies for public funding. Through the analysis of the literature (Chapter 2), it became evident that in order to increase our understanding of the factors driving HTA coverage decisions in Europe, the following strategies would be required: i) increasing the scope of analysis to multiple EU HTA bodies; ii) enhancing the comprehensiveness of the factors that are assessed; and iii) using alternative methods of analysis that would better capture those factors that are able to explain variation in coverage decisions within and between HTA bodies.

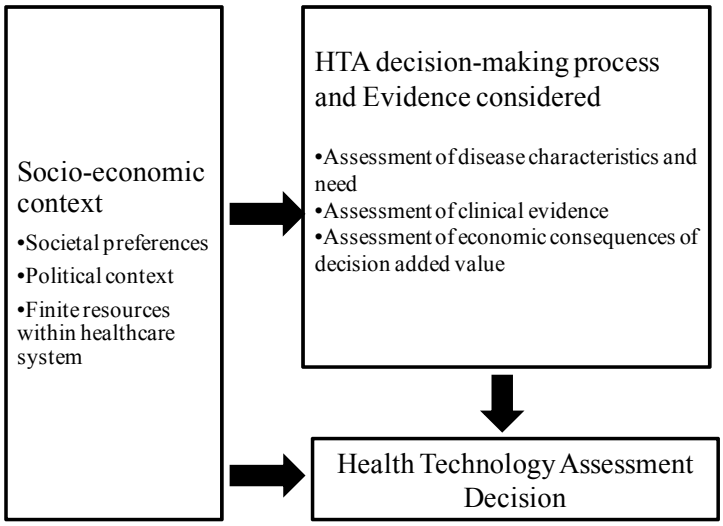
The analytical framework, described in the previous chapter, was used to guide the methods applied in this thesis. The premise of this framework is that examining characteristics of the technology, the decision process and the socio-economic context of coverage decisions will provide a comprehensive view of HTA decision-making and a platform from which to examine factors driving HTA outcomes. In line with the analytical framework, this chapter discusses the methods that were used to address the research question. In particular, this chapter presents the criteria for selecting the HTA bodies for analysis, the process of identification of variables for inclusion, the methods used to create a bespoke dataset for the analysis, and the methods of statistical analysis. Finally, the potential limitations of the proposed methods are discussed.

## 3.1 HTA decision-making: hypothesised drivers

As discussed in Chapter 2, upon examination of the literature, three "streams" of research on factors impacting on HTA decision-making became apparent. Firstly, research on the impact of evidence on the HTA decision, which focuses on the degree to which evidence related to the medicine under review (whether clinical, economic or otherwise) can impact on HTA coverage decisions (including Clement et al. 2009, Dakin et al 2006, Devlin and Parkin 2004, Mason and Drummond 2009). Secondly, research on the decision-making process itself, rather than on the technology (e.g. whether economic evaluation is a component of the decision-making process or not). The literature examining the HTA appraisal process provides insights into a number of process-related factors that can potentially influence coverage decisions (including Al et

al. 2004, Menon et al. 2005, Vuorenkoski et al. 2008, OECD 2005). Thirdly, reference in the literature is made to the impact of the overall healthcare and welfare characteristics on HTA decision-making (e.g. the impact of healthcare expenditure levels and the health policy priorities being reinforced by the Ministry of Health at the time of HTA decision-making). The literature has shown that coverage decisions are influenced by macro factors, such as healthcare spending per capita, societal willingness to pay, the structure of the healthcare system, as well as ethical and social considerations (including Packer et al. 2006; Owens 1998; Buxton 2005; Bryan et al. 2007; Ross 1995). Given these themes in the literature, it was hypothesised that HTA decisions were driven by the HTA decision-making process, the evidence considered within that process, and by the socio-economic and political context in which the decision was made (Fig. 3.1)

**Figure 3.1. HTA decision-making: hypothesised drivers**




Multinomial logistic regressions were used to assess the hypothesized influence of evidence, process and contextual variables on the different coverage outcomes from the HTA bodies (recommended, restricted or not recommended). The underlying relationship between the HTA outcomes and the 3 set of influential variables was thus modeled as:

$$Log\left(\frac{P(Y=j|E_L,P_M,C_N)}{1-P(Y=j|E_L,P_M,C_N)}\right) = \alpha_j + \sum_{l=1}^{L}\varepsilon_{jl}E_l + \sum_{m=1}^{M}\pi_{jm}P_m + \sum_{n=1}^{N}\gamma_{jn}C_n$$

Which results in:

$$P(Y = j|E_L, P_M, C_N) = \frac{e^{X_j}}{1 + e^{X_j}}$$

With $X_j = \alpha_j + \sum_{l=1}^{L} \varepsilon_{jl} E_l + \sum_{m=1}^{M} \pi_{jm} P_m + \sum_{n=1}^{N} \gamma_{jn} C_n$

And where:

| | |
|---|---|
| $P(Y = j\|E_L, P_M, C_N)$ | is the probability of belonging to group j given $E_L, P_M, C_N$ |
| $E_1, \dots, E_L$ | is the vector E of L explanatory EVIDENCE variables, |
| $P_1, \dots, P_M$ | is the vector P of M explanatory PROCESS variables, |
| $C_1, \dots, C_N$ | is the vector C of N explanatory CONTEXT variables, |
| $\varepsilon_{j1}, \dots, \varepsilon_{jL}$ | is the vector $\varepsilon$ of the L coefficients corresponding to the L explanatory EVIDENCE variables |
| $\pi_{j1}, \dots, \pi_{jM}$ | is the vector $\pi$ of the M coefficients corresponding to M explanatory PROCESS variables |
| $\gamma_{j1}, \dots, \gamma_{jN}$ | is the vector $\gamma$ of the N coefficients corresponding to N explanatory CONTEXT variables, |
| $\alpha_j$ | is the constant. |

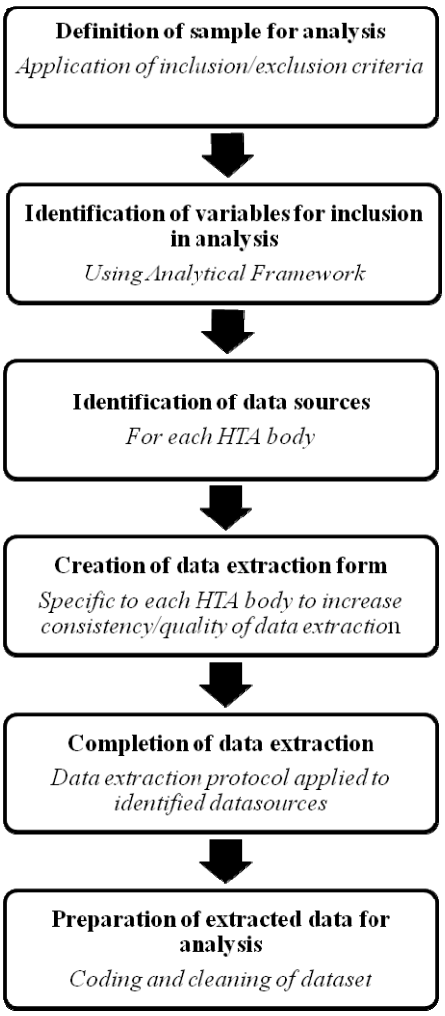
Significant efforts were made to create a dataset that was consistent and comprehensive, and able to address the research questions. It was a specific objective of this research to ensure that a comprehensive set of common variables was collected across the HTA bodies to avoid model misspecification. From the general framework characterized by the equations presented above, overall research objectives and HTA-specific objectives were derived to reflect the understanding of the context and healthcare system within which each HTA body operates. A common set of variables, complemented by HTA–specific variables where pertinent for the research objectives, were subsequently selected to enable the measurement of the role of evidence, process and context in HTA decision-making, illustrated by $E_L$ , $P_M$ and $C_N$ above. The literature that supports the selection of the underlying variables is presented in Chapter 2. The variables included in the analyses are described in Section 3.2 and in Chapters 4-8.

### 3.2 Creating the Dataset – a new and unique platform for analysis

It is recognized in the literature that HTA guidance to grant or deny funding for new technologies is influenced by different types of factors related to the evidence and disease characteristics, the process by which the decision is made, the healthcare and welfare systems, societal context and the economic environment. Based on the understanding that multiple factors are involved in coverage decision-making, from a methods point of view it was desirable to be able to explore a broad range of factors to capture as much as possible the reality of this decision-making. A need was therefore identified to develop a database of coverage decisions that captured the relevant information. A data set of information pertaining to NICE, SMC, CVZ and HAS appraisals was created, collecting information on variables relating to (i) the clinical and economic characteristics of the technology under appraisal, as well as (ii) the processes used to come to coverage decisions, and (iii) the socio-economic context in which these decisions were made.

To create this database, several steps were implemented (Figure 3.1). Inclusion and exclusion criteria were established to define the sample of appraisals for inclusion in the analysis. The variables for inclusion were defined in accordance with the analytical framework. Publicly available sources of data containing information on the variables of interest were identified for each HTA body. A data extraction form was developed and used to extract relevant data from the identified sources in a way that was transparent, reproducible and as consistent as possible between the appraisals and HTA bodies. Finally, the resulting extracted data was coded and prepared for analysis.

**Figure 3.2 Process for the development of a dataset of coverage decisions made by NICE, SMC, CVZ and HAS between 2004-2009**




### *3.2.1 Defining the sample for analysis*

The choice of HTA body included in the analysis aimed to maximize the chance of obtaining useful data to address the research question, provide a comprehensive platform for analysis of the research question rather than the examination of a particular factor in isolation, and to allow for the exploration of variation in the implementation and drivers of HTA decision-making. To gather information on variables related to the coverage decision for specific technologies, data were required from HTA bodies that published their appraisal decisions and rationale for those decisions in a comprehensive format that was accessible to the public. The following HTA bodies in Europe that fulfill these data requirements were identified:

- Scottish Medicines Consortium (SMC), Scotland (UK) National Institute for Health and Clinical Excellence (NICE), England and Wales (UK);
- College voor Zorgverzekeringen (CVZ), the Netherlands ; and
- Haute Autorité de Sante (HAS – French Health Authority), France.

The drug technology appraisals performed by NICE, SMC, CVZ and HAS formed the basis for the sample included in this analysis. The composition of the sample was determined through the following inclusion and exclusion criteria. The sample included all drug technology appraisals (as opposed to medical devices or other interventions) made in 2004-June 2009, indicated for an adult population (≥18 years). To capture a sufficient number of appraisals for both individual and aggregate analyses, a five-year time horizon was implemented. Technology appraisals were excluded from the analysis for any of the following reasons: if they focused on a non-adult population; if they appraised non-drug interventions; if marketing authorisation was withdrawn; if the ASMR was not reported (HAS only); if an abbreviated or Independent Review Panel (IRP) guidance was issued (SMC only); or if the full guidance was not available.

With regard to the HAS, additional inclusion criteria were employed. The French Haute Autorité de Sante (HAS), has numerous responsibilities, one of which is the provision of advice on new technologies available for patients. In total, the Commission issued more than 2,600 recommendations between 2004 and 2009. Given the resource constraints available, it was not possible to review all 2,600 recommendations to identify those of relevance for this research (i.e. not all recommendations provide an ASMR, some recommendations are related to a new mode of administration, new safety information or a re-review of technologies licensed prior to 2004 etc). In order to extract a relevant sample for this research, a list of technologies that had been included in the SMC and NICE reviews was created and available HAS recommendations linked to these technologies were extracted for review. While understanding that this approach may lead to selection bias, the benefit of this approach was that it increased the opportunity for comparability across agencies by collecting information on a common list of compounds, streamlining data extraction to those appraisals that were relevant for the research question.

### *3.2.2 Determining the explanatory variables for inclusion in the dataset*

In alignment with the framework set out in section 3.1, three sets of explanatory variables were considered, corresponding to the three components of analysis: (i) variables characterizing the evidence base and disease for which the technology is indicated (Component 1) (ii) factors relating to the appraisal process (Component 2) and

(iii) socio-economic variables reflecting the healthcare system and welfare system context in which coverage decisions are made (Component 3).

Explanatory variables related to evidence and disease characteristics - Component 1

The first set of explanatory variables was related to the evidence and disease characteristics considered by the decision-makers in the appraisal. The quality and quantity of data inputs, as well as the context pertaining to the disease and respective population, are hypothesized to be key variables that impact on HTA coverage decisions. A significant proportion of studies examining the factors that drive coverage decisions have considered the clinical and economic characteristics of the technology to be key drivers of decision-making (Devlin and Parkin, 2004; Dakin et al., 2006; Sculpher and Drummond, 2006; Tappenden et al. 2007; George et al. 2001; Clement et al. 2009). In addition, as recognised in the literature, the degree of uncertainty underlining the key characteristics of a technology under evaluation and its use in clinical practice can influence the probability of obtaining a positive or negative recommendation (Devlin and Parkin, 2004; Clement et al. 2009). To reflect the nature of the evidence supporting the technology, and the degree of uncertainty around the evidence, data on the quantity and quality of the clinical and economic characteristics of the technology was extracted. Clinical variables included the characteristics of the randomized clinical trials, use of non-randomised data, and the disease characteristics that the technology is licensed to treat. Several types of economic variables were considered in the analysis – the use of cost-utility analysis (CUA) was documented, and for those appraisals that included CUA, the incremental cost-effectiveness ratio (ICER) was recorded, along with measures of uncertainty around the ICER (both probabilistic and univariate measures of uncertainty). A previous analysis of NICE decision-making by Dakin et al. (2006) did not consider the impact of the degree of uncertainty around the ICER, although Devlin and Parkin (2004) and Clement et al. (2009) did include economic uncertainty as a variable in their analyses. To capture as much as possible the clinical and economic evidence supporting the technology, as well as the disease characteristics of relevance for the technology, the following variables were extracted for analysis for each appraisal:

- Characteristics of randomized clinical data
    - Number of RCTs included in the submission
    - Size of the population included in RCT(s)

    - Statistically significant results reported in RCT(s)
    - Length/extent of follow-up in the RCT (weeks)
    - Use of active or placebo comparator in RCT(s)
- Other non-randomised clinical data on efficacy/effectiveness
    - Number of observational studies included in the submission
- Disease-area / treatment paradigm
    - Prevalence of the disease/clinical condition for which technology is indicated
    - Availability of alternative pharmaceutical therapies in the current treatment setting
    - Disease area category (as per British National Formulary categories)
    - Orphan designation status
- Economic Evaluations
    - Inclusion of Cost Effectiveness Analysis in the submission
    - ICER of technology vs. comparator
    - Reporting of multiple ICERs
    - Uncertainty around the ICER reported in the submission (univariate and probabilistic)
    - Anticipated incremental pharmaceutical budgetary impact of introducing the new technology into the health care system
    - Use of societal perspective

<u>Explanatory variables related to decision-making process - Component 2</u>

Explanatory variables related to the decision-making process were also be extracted for analysis. This included a range of variables that captured key aspects of the appraisal process. Of particular interest was the impact of specific components of the decision-making process on the decision outcome – namely, the presence of a cost-effectiveness evaluation and/or a budget impact assessment (Packer et al. 2006; Devlin and Parkin 2004; Tappenden et al. 2007; Clement et al. 2009; George et al. 2001). Hutton et al. (2006) also highlight the relevance of considering the characteristics of the decision-makers, how many are involved in the decision-making and whether the HTA body is independent of the Department for Health or part of it. It was hypothesized that in the HTA appraisal process where patient group submissions are permissible, the use of such submissions may increase the probability of a recommendation (Dakin et al. 2006).

Within NICE decision-making, the use of an MTA or STA process has been hypothesised to impact on decision-making (Barbieri et al. 2009)[1]. Finally, whether the HTA decision-maker is aware of the price of the technology prior to the appraisal may also impact on the coverage decision, as opposed to a body that appraises a technology without knowledge of its price. The process-related factors that were included in the analytical framework include:

- Components of the decision-making process
    - Inclusion of patient submissions
    - Cost-effectiveness evaluation component within the process
    - Budget impact as a component of the decision-making process
- Price of the technology known at the time of appraisal
- Number of accountable decision-makers
- Accountability for the drug budget by the agency
- Independence of the decision-making agency
- Number of technologies appraised simultaneously

Explanatory variables related to the socio-economic context of decision-making - Component 3

Finally, macro-economic, societal and healthcare system-related variables were included to reflect the context in which coverage decisions are made, and to assess their degree of impact on HTA decision-making. There is less variation between appraisals with regard to these indicators because they are defined at the country level rather than at the appraisal level. For example, it was hypothesized that the percentage of GDP spent on healthcare, and the percentage allocated to the drug budget, may impact on the rate of reimbursement of new technologies (OECD 2005; Packer et al. 2006). It was also envisaged that government prioritization of specific disease areas may help spur reimbursement for technologies emerging in those priority disease areas, potentially at the expense of other technologies indicated for 'less-priority' diseases. With regard to the year of appraisal, HTA appraisal processes can/have changed over time and this may impact on the coverage decision. Population size, in terms of the population under the

[1] Multi technology assessments (MTA) and Single technology assessments (STA) are two appraisal processes used by NICE to issue guidance on technologies. The MTA process includes a third-party review and analysis of the clinical and economic evidence, while the STA process relies on manufacturer provided information (Barbieri et al. 2009).

agency's remit, was considered relevant, as it can have implications for the budgetary impact of new technologies, absolute levels of funding that are available, and absolute levels of demand. The extent to which decisions for funding are made exclusively at the national level, or whether the agency in question operates at a regional or local level were also considered as pertinent factors to record. Finally, given that healthcare policy is influenced by the political context in which it is made, it was thought relevant to capture whether decisions were being made in an election year. To this end, the following variables were captured:

- Date of submission
- Population size – agency coverage
- GDP-healthcare expenditure (as percentage of GDP spent on healthcare)
- Healthcare expenditure on pharmaceuticals per patient per annum
- Drug funding process within the healthcare system – whether centralized or decentralized
- Election year at the time of the decision
- If technology was indicated for a priority disease area

### *3.2.3 Data sources*

Data sources for the analyses of coverage decisions were organized by the three components of analysis. Data sources related to the clinical and economic evidence as well as the disease relevant to the technology (Component 1 of the analytical framework) were derived from the reports of coverage decisions by NICE, SMC, CVZ and HAS. Data sources related to HTA appraisal processes (Component 2) were derived from the agencies themselves, as well as from national Department of Health sources and the literature. Finally, data sources to capture macro variables related to the healthcare system, as well as societal and welfare system characteristics, were captured from ministry of health documents, national statistics publications, and published literature.

To gather information on variables related to the characteristics of the technology under appraisal (Component 1), data were obtained directly from each of the HTA web portals. For NICE, information on appraisals was collected from several types of sources available on the NICE website. The Final Appraisal Determination (FAD) was used as the key source of data for extraction. In addition, Assessment group, Expert

Review Group reports, as well as stakeholder submissions and manufacturer reports were also accessed to extract needed information. For the SMC, the appraisal report for each technology was accessed. For the CVZ, a series of documents were used as sources of data for the creation of the dataset. This included the official letter from the CVZ to the Ministry of Health, providing information on the final recommendation of the CVZ. Additional sources include the CVZ report, and in addition the cost consequence report, where applicable. For technologies applying for inclusion in the 'expensive drug list' the particular document created by the CVZ which documents the available information and intended analyses to be performed in the future was utilised. Finally, to ascertain the coverage decisions, the following documents were accessed: the official 'expensive drug list', the Medicijn Kosten web-portal (which documents all medicines available in the Netherlands and gives information on whether there is a co-payment for the patient, a restriction on use etc); and the Farmaceutisch Kompas which lists all of the technologies reimbursed in the medicine reimbursement system (*Geneesmiddelen Vergoedings Systeem*, GVS) under either List 1A, 1B or List 2 (See Section 6.2). To gather information on the clinical and disease characteristics of the technologies appraised by HAS, the Transparency Committee's reports were accessed.

With regard to data on payer decision-making processes (Component 2), a range of data sources were examined to extract information about the characteristics of the appraisal process. HTA guidelines for manufacturers and the methodology guidelines were accessed from the NICE and SMC websites. For the SMC, the minutes of the Committee meeting in which the appraisal was discussed was also accessed from the SMC website, to ascertain the members that were involved in the decision-making process as well as to collect any insight on other factors or rationales taken into account that were reflected in the minutes. Legal documents outlining the appraisal process were utilised for both the CVZ and HAS. In addition, information about the appraisal process was extracted from the literature for each of the HTA bodies (e.g. Mason and Drummond 2009; Sorenson et al. 2008; Clement et al. 2009; Schäfer et al. 2010; Sandier et al. 2004).

Finally, with regard to data related to the context in which payer decisions are made (Component 3), multiple sources were utilised. Information about expenditure on healthcare for each for the British, French and Dutch healthcare systems were extracted

from the OECD database and from Ministry of Health figures. The European Statistics and Population database, as well as specific population statistics from the respective jurisdictions were utilised to estimate total national population and the population within the remit of the agency. Government and media publications were accessed to identify years when national elections took place, and what the priority disease areas were for the respective healthcare systems.

The detailed sources of data used for each HTA body and variables are provided in the description of the dataset in the empirical chapters that follow (Chapters 4-7).

### *3.2.4 Data extraction*

To encourage consistency and rigour in the data extraction process, a data extraction protocol was developed. This ensured, to as great a degree as possible, that the relevant data was extracted from the identified sources in a way that was transparent, reproducible and as consistent as possible among the different appraisals and HTA bodies.

The data extraction form was organized into three segments, relating to the three components of analysis that are integral to this research. The first segment was designed to allow for extraction of information relating to the decision itself – the nature of the clinical evidence available, whether cost-effectiveness evidence was put forward, and if so, the characteristics of that evidence. The second segment captured information relating to the process by which the recommendation was issued – for example, within NICE appraisals, there is the possibility to appraise via a Multi Technology Appraisal (MTA) or a Single Technology Appraisal (STA). Finally, the third segment of the data extraction form captured information on the context in which the guidance was issued (healthcare system, economic and social context).

## 3.3 Statistical Analysis

Having set the scene by providing both the list of dependent and explanatory variables for analysis, as well as the HTA agencies which were analysed, this section outlines the statistical methods that were employed to analyse the data set and to examine the importance of a range of diverse factors on coverage decisions made by NICE, SMC, CVZ and HAS from 2004 to 2009.

### *3.3.1 Descriptive statistics*

Descriptive statistics were calculated for each HTA body for each extracted variable, stratified by outcome group (recommended, restricted, or not recommended). This included the mean, standard deviation, median, minimum-maximum range observed, and confidence interval. For categorical variables the chi-squared test was performed across the three outcomes to ascertain if a statistically significant difference of the means between recommended and restricted technologies as well as between recommended and non-recommended technologies was present, at a level of significance of 0.05.

With regard to non-categorical variables, to determine the relevant statistical tests to use in assessing the significance of differences observed between means (at 0.05 level), it was necessary to assess whether the normality assumption was valid for the variables under consideration. For all variables, the sample is >30. It has been suggested that when analysing sample sizes of >30, even when the normality assumption is violated, parametric tests may still be performed (Pallant 2007). Prior to making a decision on which statistical tests to apply to specific variables, the distribution for each variable was further examined.

To test the normality assumption the following was performed for each non-categorical variable: i) skewness and Kurtosis was calculated, and ii) the standardised normal probability plot (P-P plot) was graphed. The results of the above-described tests for each HTA body are presented in Appendices A, B, C, D for NICE, SMC, CFH and HAS respectively. Given the results of the tests, it was considered appropriate to run both a range of statistical tests to ascertain the presence of a statistically significant difference between the observed means for recommended versus restricted technologies as well as between recommended and non recommended technologies. Two parametric tests were performed: i) Student's T-test and ii) one-way ANOVA. For the t-test, first the means of recommended versus restricted interventions were compared, then restricted versus not recommended were compared via a Student's unpaired T-test. Bartlett's test was performed to ascertain if the t-tests should assume equal or unequal variance. The Kruskal-Wallis test was also implemented. Statistical analyses were conducted using Intercooled (IC) STATA (Version 10.1 2009).

### *3.3.2 Multivariate analysis – considerations*

A number of considerations were taken into account in selecting the statistical model that would be most appropriate for the analysis, including the definition of the outcome variable, the ordinality of the data. Methods for characterising and handling of missing data were also considered. In developing and performing multivariate analysis, the objective was to identify the individual effect of each of the factors considered, controlling for their potential association with the rest of the factors in the model, and to assess which combination of factors best explains the pattern of coverage decisions observed within each HTA body, as well as across HTA bodies. Multivariate analyses of each HTA agency were undertaken to understand the specificities of each HTA agency in the factors that drove coverage decisions within the appraisal process. Pooled analyses of the data in the four HTA bodies were then conducted to better understand the sources of variation in coverage decisions between HTA agencies and jurisdictions. Thus, the combination of single and pooled analyses was used to address the research question in order to explore single agency decision-making, and to test for differences on the impact of factors on cross-agency variation in coverage decisions. The face validity of the multivariate analysis results was assessed by conducting interviews with representatives of NICE, SMC and CVZ in which the model results and interpretation of results was discussed and assessed. A representative of the HAS as not available for comment. Interview summaries are provided in Appendix E.

<u>Categorisation of the outcome variable and implications for analysis</u>

Econometric analyses of coverage decisions have adopted both binary and multi-category outcome variables as the basis for their analysis (e.g. Dakin et al. 2006; Devlin and Parkin 2004; George et al. 2001; Clement et al. 2009). For example, Devlin and Parkin (2004) adopted a binary outcome variable, thus modelling the impact of five explanatory variables on the odds of a recommendation or a non-recommendation by NICE. Similarly, George et al. (2001) adopted a list/not list binary outcome variable in their analysis of Pharmaceutical Benefits Advisory Board (PBAC) coverage decisions in Australia. In contrast, both Dakin et al. (2006) and Clement et al. (2009) adopted a three-category explanatory variable. Dakin et al. (2006) adopted a three-category outcome variable, examining the impact of explanatory variables on the decision between recommendation, restriction and non-recommendation by NICE. Similarly,

Clement et al. (2009) considered 'list', 'list with conditions', and 'do not list' categories in their analysis of PBAC, Canadian Drug Review (CDR) and NICE coverage decisions, although they did not perform a regression analysis of their dataset, but focused on a descriptive analysis of the data by examining the role of each variable separately from the other. Therefore, there are examples from the literature of the use of both binary and three-category outcome variables to analyse coverage decisions.

In parallel to a review of outcome variables used in the literature on coverage decisions, an assessment was also made of how each HTA body operates in terms of its decision-making. All four agencies appear to consider more than two types of coverage options when making their final recommendation. HAS uses a five point scale known as the ASMR rating, that classifies the technology according to the level of incremental medical service rendered, with the highest level - I - being high incremental medical service and V denoting no additional medical benefit. Subsequent pricing and volume negotiations are based on these ratings. Thus, a binary approach would not fit the way in which HAS approaches coverage decisions. Similarly, the CVZ has several possible coverage options in its armamentarium. According to whether the technology is for inpatient or outpatient use, it can cover a technology under a special 'expensive drug list', in List 1A, where price is clustered to the lowest price in the cluster, List 1B for those technologies that cannot be clustered or include a technology in List 2, which is for those technologies (either in 1A or 1B) that have additional conditions attached to them: e.g. use in a specific patient group etc). Thus, more than two coverage options are also available to the CVZ committee. The SMC utilises the three-category approach: accepted for use, accepted for restricted use and not accepted for use in NHS Scotland, therefore clearly not operating in a binary-type coverage decision system. Similar to the SMC, NICE either recommends, recommends with conditions or does not recommend technologies for NHS coverage in England and Wales. In all four HTA bodies therefore, coverage decisions are based on more than a binary decision-process. This was therefore a key motivation for justifying a non-binary modelling approach. The outcome variable for this analysis was categorised into three possible outcomes, where the new technology can be

- recommended for routine use
- recommended for restricted use

or

- not recommended

The choice of a three-category outcome variable was selected to best reflect the way in which HTA bodies consider their coverage options. This choice was done with full knowledge of the complications associated with using a three-category dependent variable from a statistical modelling point of view, but more importantly due to complications arising from the heterogeneity in the types of coverage decisions/types possible – particularly with regards to the 'restricted' category. This heterogeneity has been recognised both by Raftery (2006) as well as O'Neill and Devlin (2010). Raftery (2006) identified various forms of restrictions made by NICE in its appraisals, which were categorized into 'major' and 'minor' restrictions. Major restrictions included, for instance, the recommendation of the technology for a specific sub-population of the licensed indication, or recommendation of the technology for use in second or third line use. Minor restrictions included the requirement to monitor as a condition for using the technology or prescription being restricted to a specialist. A more quantitative assessment of the heterogeneity of restrictions made by NICE was performed by O'Neill et al. (2010). In this assessment, restriction was evaluated by calculating 'M' or the proportion of patients for whom the technology was recommended, relative to the population that would be eligible for the technology as per the license:

$$M = (p/P)*100$$

where M is the level of patient access (0=0%, 1 = 100%), P is the total eligible population and p is the recommended population within the coverage decision (O'Neill and Devlin 2010).

The results of this quantitative analysis of NICE restrictions (n=34) showed that the degree of restriction varied between technologies/appraisals: 71% of technologies had an M of 50 or less, 32% had an M of 10 or less, while approximately 30% had an M of 50 or more. As a more precise measure of the impact of coverage decisions on patient volume, the 'M' concept usefully illustrated the variability in the level of access associated with a restriction. However, the use of 'M' as an outcome variable for the analysis of coverage decisions is limited by (i) non-availability of the data to run this calculation; (ii) its focus on a specific type of restriction; and (iii) by its assumption that NICE is aware of the impact of its restriction on the size of patient access prior to its recommendation. In approximately half of the appraisals considered by O'Neill and

Devlin (2010) it was not possible to calculate M, due to poor information on the total population eligible and the estimate of the population to which the technology was recommended. While M calculates the impact of a restriction on patient volume, it does not take into consideration other types of restrictions that can be issued by NICE, such as restrictions on specific types of prescribers, with specific monitoring (Raftery 2006). Finally, the analysis is also limited by its assumption that NICE, when making its restriction, takes into account the impact of that restriction on patient volume. However, there is no evidence to suggest that this is the case: usually, target population size is calculated as part of costing templates, and these templates are produced as a result of the NICE appraisal conclusions, rather than as a factor influencing the outcome of the appraisal. While there are differences in the approach to analyzing 'restricted' technologies, both Raftery (2006) and O'Neill and Devlin (2010) suggest that the impact of the restriction can vary and that the use of restrictions is heterogeneous. While this variability in the 'restricted' category has implications for the interpretation of the results of multivariate analyses, the notion of removing this restricted category and considering a binary approach does not provide a ready solution, as it would simply lead to a category in which recommended and restricted coverage decisions are combined, increasing the heterogeneity of the category even further and reducing the ability of the analysis to reflect the way HTA bodies perceive what the coverage options available to them are.

In summary, given that this thesis aims to analyse the factors that drive HTA coverage decisions, the analysis was designed to consider coverage decisions in the way that HTA bodies define them (i.e. in multiple options rather than a binary (yes/no) option). From this perspective, a multinomial logistic approach was deemed attractive as it is an appropriate method for handling outcome variables with multiple non-ordinal categories. Using a three-category outcome variable reduced the need to collapse the multiple variable categories into binary categories, and thus better reflected the multiple coverage options available to HTA bodies. However, to ascertain the impact of model characteristics and assumptions on understanding HTA decision-making, the results of the base case models for both single HTA body analyses as well as pooled analyses, were compared with models specified with binary outcome categories (see below).

Ordinality of the outcome variable

In addition to defining the number of categories for inclusion in the outcome variable, an important consideration from a methodological standpoint is whether ordinality should be assumed, as this has implications for the statistical methods that should be applied. A categorical variable can be considered ordinal if there is a 'natural' ordering in the outcome that can be identified. Dakin et al. (2006), in their econometric analysis of NICE decision-making, used a three-category approach, and argued that an assumption of ordinality was not appropriate as there was no consensus on the 'direction' of the ordering. In other words, depending on whether the aim of the appraisal decision was to define volume (i.e. number of patients that can access the therapy), total drug cost for the technology (budget impact) or maximization of patient/healthcare system outcome can lead to different ordering and consideration of appraisal decisions. For example, if the main objective of the coverage decision was to control cost, then it could be expected that those decisions which restrict or not recommend access may be considered as preferable to a recommendation. The opposite is true if we consider patient access to medications as the main objective of coverage decisions: in this case, a coverage decision that covers 100% of the eligible population, less than 100% or 0% could be considered. In the base case analysis, multinomial logit regression was assumed to be appropriate, although sensitivity analyses in which ordinality was assumed were run for each HTA body and for the pooled data set to examine the impact of that assumption on model results (see below).

Missing data and imputation

While every effort was made during the data extraction process to identify the information relevant to the variables of interest, a proportion of data was missing. Missing data can impact on the results of the analysis if conclusions are drawn from biased parameter estimates, due to reliance on the complete observations within the dataset or to biases arising in the imputation of missing explanatory variables. To generate robust multivariate analyses of HTA decision-making, it was important to identifying the most suitable means of handling missing entries in this data set.

To assess the extent to which imputation of missing data would be appropriate, the rate and pattern of missing data for each HTA body was first examined, alongside an assessment of the process by which data was missing and whether an assumption of

"missing at random" (MAR) was justified (Rubin 1976, Lu and Copas 2004). The proportion of missing entries for each HTA body within the dataset was relatively small and ranged from 5% to 10% (Table 3.1). On average, for each appraisal, no more than three indicators were missing, of the 30+ collected variables. The extent of missing information across the outcome categories was also examined and showed that the proportion of missing values were similarly distributed across the outcome variable, suggesting that the impact of any biased imputation would not fall disproportionately on a particular outcome category (Table 3.1).

With the exception of the HAS, economic-related variables capturing information on the uncertainty estimates around the ICER suffered from the highest proportion of missing entries, driven from lack of reporting of this information in the public domain (Table 3.1). Missing values for these variables were imputed using information from other variables, particularly the ICER itself, and a set of clinical trial related variables. However, the imputed uncertainty variables were finally excluded from the models due to the extent of missing data and because of the high collinearity with the ICER explanatory variable.

**Table 3.1 Description of missing data, by HTA body**

| | NICE | SMC | CVZ | HAS |
|---|---|---|---|---|
| Proportion of incomplete entries (%) | 8% | 10% | 9% | 5% |
| Missing entries by outcome variable | Recommended: 5%<br>Restricted: 9%<br>Not recommended: 9% | Recommended: 10%<br>Restricted: 9%<br>Not recommended: 10% | Recommended: 9%<br>Restricted: 8%<br>Not recommended: 9% | ASMR 1-2: 7%<br>ASMR 3-4: 4%<br>ASMR 5: 4% |
| List variables with missing values (top three variables) | Univariate uncertainty estimates around ICER<br>Probabilistic sensitivity analysis | Univariate uncertainty estimates around ICER<br>Probabilistic sensitivity analysis | ICER<br>ICER –related sensitivity analysis | Size of RCT<br>Duration of RCT<br>Demonstration of clinical superiority |
| Missing entries per appraisal (mean and range) | 3 (range: 0-9) | 3 (0-10) | 3 (range: 0-8) | 1 (range: 0-8) |

A relatively low proportion of missing values were noted for other variables (Appendix A-D Figures A.1, B.1, C.1, D.1, provide data on the distribution of missing entries per variable for each HTA body). Within the HAS appraisals, clinical variables

including the size, duration and statistical significance of the results were variables with the highest number of missing entries. The likelihood of missingness for these indicators was mostly linked to the reason for appraisal and the nature of the technology. Specifically, amongst the appraisals with missing RCT-related variables, 56% were re-appraisals of previous HAS decisions and 17% were orphan designated medicines. The mean ASMR among appraisals with missing RCT-related entries was 3.5 compared with 3.9 for those appraisals with complete RCT-related entries. To consider these variables in the model, the imputation process utilized the comprehensive range of variables collected within the data extraction process. For example, in the instance that a particular clinical variable was missing, such as RCT sample size, the size the eligible population for treatment and therapeutic area (BNF category), coupled with additional clinical variables (use of active comparator, number of clinical trials) were identified as variables that could be utilized in the imputation process. A more detailed description of the imputation approach taken is provided below.

To maximize the number of observations and sample size for both the individual and pooled analyses, imputation techniques were used to estimate entries for those observations that were lacking. In order to determine the method most adapted to the analyses, several approaches to imputation were tested including (i) imputation by replacing missing values with the overall mean of the variable, (ii) generating regression estimates of the missing value, and (iii) multiple imputation techniques that take random variation into account. This exercise was performed on one of the HTA datasets (SMC dataset) to test which imputation method would be most useful to extrapolate to the remaining HTA multivariate analyses. The results of this exercise are summarised in Appendix F. The results suggest that the various imputation techniques provided similar pseudo R-squared results across the models and similar pattern of size, direction and significance of effect. It was therefore felt to be appropriate, for the remainder of the multivariate analyses to impute missing values by using regression estimates of the missing value (option ii, performed using the 'impute' command in STATA data analysis software (Intercooled (IC) Stata version 10.1). For each imputed dataset, imputation output was checked for coherence. In addition, dummy variables were created for those factors that had observations with incomplete entries, so as to be able to ascertain, through regression analysis, whether the lack of data was significantly associated with the outcome variable. Sensitivity analyses were conducted in which

those observations with missing entries were removed from the models so as to allow for an assessment of the impact of imputation (or no imputation) on the model results. This sensitivity analysis was implemented with the knowledge that removing those observations with missing data could lead to selection bias in the analysis due to the fact that appraisals with incomplete observations may be different than appraisals with complete observations. However, what this sensitivity analysis wished to assess was the implications of making a trade-off between including in the analysis only those observations with data, therefore avoiding imputation but potentially increasing bias in the analysis, or maximising sample size and validity of the model output by utilising imputation techniques.

### *3.3.3 Multivariate analyses*

Model specification

As indicated in Chapter 3, a large number of interrelated factors could be hypothesised to influence the coverage decisions taken by HTA bodies. The analysis has collected data relating to many of these factors (see Chapter 3, section 3.1.2). In terms of the statistical modeling of the coverage decisions, however, a process needed to be developed in order to select the set of explanatory variables that appeared in the final specification of each of the models estimated. This selection process involved the following steps:

1. The referent outcome variable was defined. The base outcome utilised for all four multinomial regression models was the 'recommended' outcome. Thus, recommended technologies were compared with restricted and non-recommended technologies each in turn.
2. The results of the descriptive analyses of the dataset were reviewed
3. The full set of variables extracted for each HTA body was considered in the model specification process. This was done by examining the impact of potential explanatory variables in 'mini' bi-variate regression models. In addition, dummy variables were created for incomplete observations, to be able to explore their impact in the model.
4. Those variables with a p value of $\leq 0.25$ were included in the preliminary model. Subsequently, the coefficients and significance of the variables in the preliminary model were compared with the coefficients and significance observed in the bi-variate models and any evidence of interaction was noted.

5. Using the preliminary model as a basis, the model was reduced by removing those variables that were significantly above the 0.10 threshold. The coefficients and significance of the variables remaining in this 'reduced' model were compared with the coefficients and significance levels observed in the preliminary and bi-variate models to seek evidence of interaction effects.
6. To examine the stability of the model, re-estimation of the 'reduced' model was performed by sequentially removing one variable at a time and observing the effect on the coefficient and significance level of the remaining variables to the original 'reduced' model. If the variables remained stable in their effect size and significance, this was considered the base case model.
7. This base-case model was subsequently tested through alternative model specifications (further described below).
8. As a final step, the base-case model results were presented to representatives of the HTA body to which the analysis pertained to seek feedback on the variables identified within the base case model, the coefficient and level of significance to assess the validity of the model.

By selecting only the subset of indicators with the strongest effect on the coverage decision, the systematic application of steps 1 to 8 facilitated the interpretation of the results of the models, whilst allowing for the testing of the potential impact of the wide range of indicators collected in the study.

In the first instance, multivariate analyses were conducted separately for each HTA body, to maximise the use of available data, and to allow for each analysis to be conducted specifically to match as closely as possible the characteristics of their decision-making process in terms of the process adopted, as well as to accommodate differences in the outcome and independent variables considered. For instance, HAS synthesizes its analysis of the clinical value of a technology by awarding a rating (ASMR rating) – this is a particular characteristic of the HAS system that is not observed in NICE, SMC or CVZ appraisal processes. For each HTA body, the characteristics of the sample, the protocols for method extraction and the data sources are described in the chapters that follow (4-7). In addition to modelling coverage decision-making for each agency separately, a pooled analysis of all coverage decisions across HTA bodies was identified as an important additional analysis of direct relevance

to the research question. The analysis was performed using the STATA data analysis software (Intercooled (IC) Stata version 10.1).

Alternative model specifications - sensitivity analyses

To examine the robustness of the base case model, alternative model specifications were tested in sensitivity analyses (Figure 3.2). The general aim of these sensitivity analyses was to assess to what degree the base case model results changed if model specifications for particular assumptions were altered. A sensitivity analysis was conducted examining the impact of a binary rather than a three-category outcome variable. As indicated above, the rationale for this analysis was linked to the fact that previous analyses published in the literature on NICE and SMC decision making used a binary approach in the regression analyses (Devlin and Parkin 2004; Clement et al. 2009). This was done by considering a 'covered' and 'not covered' approach: the recommended and restricted categories were grouped together, and the not recommended category was kept as in the base case analysis. Logistic regression was performed examining the log likelihood and the odds of coverage versus no coverage. Related to this, a specific sensitivity analysis was performed for the HAS sample, testing the impact of an alternative classification of HAS coverage decisions using different ASMR categories.

To test the impact of imputation on the model results, a sensitivity analysis was conducted restricting the base case analysis to complete observations only (thus excluding observations with imputed values). This sensitivity analysis was implemented with the knowledge that removing those observations with missing data could lead to selection bias in the analysis due to the fact that appraisals with incomplete observations may be different than appraisals with complete observations. However, what this sensitivity analysis wished to assess was the implications of making a trade-off between including in the analysis only those observations with data, therefore avoiding imputation but potentially increasing bias in the analysis, or maximising sample size and validity of the model output by utilising imputation techniques.

To examine the impact of assuming a non-ordinal outcome variable, a sensitivity analysis was conducted assuming the outcome variable was ordinal. The proportional

odds model, and when the assumption of proportional odds was violated, the generalized ordered logistic model were used. Further description of the methods considered for this sensitivity analysis are summarised in Appendix G.

In the base case analysis, all technologies were included in the analysis, and the disease areas to which they were related (defined by BNF criteria) was used to adjust for differences in technologies assessed between HTA bodies. In addition, the analysis was repeated using a sub-set of technologies reviewed by all four agencies, albeit at the cost of reducing the sample size, to standardise the baseline sample used in the multivariate analysis. Furthermore, the analysis was performed in which the sample was restricted to cancer therapies, as an alternative means of standardising the type of technologies assessed by each HTA body.

## 3.4 Discussion

The methods proposed for this thesis involved the creation of a specific dataset containing detailed information on the characteristics of coverage decisions by four European HTA bodies, and the quantitative analysis of this dataset using multivariate techniques. The methods were selected in the belief that they would provide informative and helpful analyses and observations to address the research question. At the same time, it is recognised that such an approach faces a number of important limitations which need to be identified and if possible, addressed.

The selection of HTA agencies included in the analysis was in part limited by data availability. In particular, only those HTA agencies that published their decision-making process and reports in the public domain were considered for analysis. Therefore, generalisation of results, even within the respective agencies, must be attempted cautiously. HTA agencies are continuously evolving and factors that may have driven their decision-making during 2004-2009 may not be the same or remain constant over time. Within the United Kingdom, the new Conservative-Liberal Democrat government elected in May 2010 has indicated that they will change NICE's role such that its reimbursement recommendations will only have advisory status. While this change in NICE’s role may or may not impact on the factors driving its decision-making, these changes would appear to impact most significantly on the

implementation of NICE guidance by the health services in England and Wales (further discussed in Chapter 4 and 9, section 9.5.2).

While every effort was made to create a dataset and analysis that was comprehensive in its inclusion of explanatory variables, invariably it was not feasible to capture all variables presumed to have a potential impact on coverage decisions. Although the analysis captured information on the ICER and the uncertainty around it, these indicators do not provide information on the nature of the economic model, the design, comparators and subtleties of the analysis submitted to the HTA body. Moreover, the use of surrogate outcomes was not captured as a specific variable in the dataset - there is increasing interest in understanding better the use of surrogate endpoints in decision-making. As the literature does suggest that most clinical trials utilise surrogate endpoints, it was considered not practical to include this as a variable due to potential lack of variability between technologies, and other variables linked to the clinical characteristics of the technology were prioritised instead (e.g. use of active comparator in trial, demonstration of superiority in clinical trial etc).

The speed of the appraisal process was also not factored in the analysis. The fact that the SMC takes 3 months and NICE takes 12 months to complete an appraisal could have an impact on the nature of the evidence considered and the impact of the process on outcomes. The analysis is also limited in that it cannot systematically examine time-to-coverage decisions as a specific outcome of analysis. Finally, the data base excludes those appraisals for which no documentation was available. This includes situations where manufacturers did not make a submission, and therefore no appraisal was conducted. This could be considered as a source of bias in the sample in that it may not consider the characteristics of those technologies that were not appraised. The proportion of such cases excluded due to non-submission were recorded and reported.

Socio-economic indicators, and particularly GDP, are correlated with many different factors. Indeed, such indicators act as a surrogate for many characteristics of the country it applies to. In addition, such indicators that vary at the HTA body level, rather than the technology appraisal level, are unlikely to have a very strong effect, due to the limited number of HTA bodies in this analysis. Therefore, the interpretation of the impact of such broad indicators, such as the percentage of GDP spent on healthcare, will

need to take into consideration the risk that variations observed in such indicators across HTA bodies may be correlated with other factors.

Finally, it is important to consider the research findings within the context of the respective origins and objectives of NICE, SMC, CVZ and HAS. The evolution of the HTA bodies, their origins and their role within pharmaceutical regulation and coverage decisions vary. In particular, their role within their respective healthcare systems is very much driven by the context and characteristics of the healthcare system, and each HTA agency has peculiarities about their role that are linked directly to that system. HTA bodies, both within and outside of Europe, vary in their objectives, and in the approach and methods used to implement HTA within their jurisdictions (Neumann et al. 2010). The scope of this thesis was limited to national level HTA bodies. In addition, the analysis of factors influencing coverage decisions were performed on decisions made within HTA processes. However, within a healthcare system, where funding/reimbursement decisions are confirmed at a national level, regional and local payers may re-assess whether funding/reimbursement should be provided. Due to time constraints, the nature of factors influencing regional/local funding/reimbursement decisions was not assessed.

Before embarking on a pooled analysis of coverage decisions across HTA bodies, the pros and cons of such an analysis were assessed. Several considerations favoured a pooled analysis: firstly, it could provide data to help explain variation in coverage decisions across HTA bodies, which is of direct relevance to the research question. Secondly, the pooling together of appraisals from four HTA agencies could significantly increase the sample size, creating one of the largest single sets of data on HTA coverage decisions and accompanying appraisal characteristics. Pooling across HTA bodies was felt to be feasible due to the fact that the data set was created specifically for this research question, and all data were extracted by the same researcher with a specific data extraction protocol. Challenges identified in performing a pooled analysis across HTA bodies were: i) HTA bodies do not formulate coverage decisions in the same way; ii) heterogeneity in the nature of technologies assessed by the different HTA bodies used in the analysis; and iii) heterogeneity in the information available for extraction. The need to address as much as possible these

limitations was a key driver in how the sample was pooled, the analyses performed, and in the choice of sensitivity analyses that were run.

In summary, the methods used to address the research question focused on three elements: i) increasing the scope of analysis to multiple EU HTA bodies; ii) enhancing the comprehensiveness of the factors assessed; and iii) using analysis methods that capture better variations in coverage decisions within and between HTA bodies. Multivariate analyses were conducted separately for each HTA body are presented in Chapters 4-7. A pooled analysis of coverage decisions across HTA bodies is presented in Chapter 8.

## 3.5 References

Barbieri, M., N. Hawkins, and M. Sculpher. 2009. Who does the numbers? The role of third-party technology assessment to inform health systems' decision-making about the funding of health technologies. *Value Health* 12 (2):193-201.

Clement, F. M., A. Harris, J. J. Li, K. Yong, K. M. Lee, and B. J. Manns. 2009. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 302 (13):1437-43.

Dakin, HA , NJ Devlin, and IAO Odeyemi. 2006. “Yes”, “No” or “Yes, but”? Multinomial modelling of NICE decision-making. *Health Policy* 77:352-367.

Devlin, N., and D. Parkin. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 13 (5):437-52.

Fenwick, E., and S. Byford. 2005. A guide to cost-effectiveness acceptability curves. *Br J Psychiatry* 187:106-8.

George, B, A Harris, and A Mitchell. 2001. Cost-Effectiveness Analysis and the Consistency of Decision Making. *Pharmacoeconomics* 19 (11):1103-1109.

Hutton, J, C McGrath, JM Frybourg, M Tremblay, E Bramley-Harker, and C Henshall. 2006. Framework for describing and classifying decision-making systems using thechnology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal of Technology Assessment in Health Care* 22 (1):10-18.

Lu, G., and J.B, Copas, 2004. Missing at random, likelihood ingnorability and model completeness. *The Annals of Statistics* 32(2): 754-765.

Mason, A. R., and M. F. Drummond. 2009. Public funding of new cancer drugs: Is NICE getting nastier? *Eur J Cancer* 45 (7):1188-92.

Neumann, P.J., Drummond, M.F., Jönsson, B., Luce, B.R., Schwartz, J.S., Siebert, U., Sullivan, S. 2010. Are Key Principles for improved health technology assessment supported and used by health technology assessment organizations? *International Journal of Technology Assessment in Health Care* 26(1): 71-78.

OECD. 2005. *Health Technologies and Decision Making*. The OECD Health Project, OECD Publishing.

O'Neill, P., and N. J. Devlin. 2010. An analysis of NICE's 'restricted' (or 'optimized') decisions. *Pharmacoeconomics* 28 (11):987-93.

Packer, C., S. Simpson, and A. Stevens. 2006. International diffusion of new health technologies: a ten-country analysis of six health technologies. *Int J Technol Assess Health Care* 22 (4):419-28.

Pallant, J. 2007. *SPSS survival manual (3rd edition)*. Maidenhead, UK: McGraw-Hill Education

Sandier, S., Paris, V., and Polton, D. 2004. Health Care Systems in Transition – France 2004. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies, 2004.

Schäfer, W., Kroneman, M., Boerma, W., van den Berg, M., Westert, G., Devillé, W., and van Ginneken, E. 2010. Health Care Systems in Transition - The Netherlands 2010. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies, 2010.

Sculpher, M, and MF Drummond. 2006. Analysis Sans Frontieres - Can We Ever Make Economic Evaluations Generalisable Across Jurisdictions? *Pharmacoeconomics* 24 (11):1087-1099.

Sorenson, C., Drummond, M., and Kanavos, P. 2008. Ensuring value for Money in Health Care: the role of HTA in the European Union. Cornwall: World Health Organization 2008, on behalf of the European Observatory on Health Systems and Policies.

STATA 10.1 (2009). STATA. College Station, Texas, StataCorp.

Raftery, J. 2006. Review of NICE's recommendations, 1999-2005. *BMJ* 332 (7552):1266-8.

Rubin, D. B. 1976. Inference and missing data (with discussion). *Biometrika* 63: 581–592.

Tappenden, P., J. Brazier, J. Ratcliffe, and J. Chilcott. 2007. A stated preference binary choice experiment to explore NICE decision making. *Pharmacoeconomics* 25 (8):685-93.

# 4 Empirical analysis of NICE coverage decisions

Adalimumab was reviewed by NICE, SMC, CVZ and HAS for the treatment of ankylosing spondylitis[1]. NICE provided the following guidance:

> 1.1 Adalimumab or etanercept are recommended as treatment options for adults with severe active ankylosing spondylitis only if all of the following criteria are fulfilled.
>
> The patient's disease satisfies the modified New York criteria for diagnosis of ankylosing spondylitis.
>
> There is confirmation of sustained active spinal disease (...)
>
> Conventional treatment with two or more non-steroidal anti-inflammatory pharmaceuticals taken sequentially at maximum tolerated or recommended dosage for 4 weeks has failed to control symptoms. (...)
>
> 1.3 It is recommended that the response to adalimumab or etanercept treatment should be assessed 12 weeks after treatment is initiated, and that treatment should be only continued in the presence of an adequate response as defined in section 1.4." (NICE 2008a section 1.1.-1.3)

This extract summarises NICE's guidance on adalimumab in which its use is restricted to the sub-set of the eligible population that have failed to respond to at least two conventional therapies. The restriction on patient population is also accompanied by a so called 'stopping rule', in which clinical assessment should be made every 12 weeks to decide whether to continue with treatment or stop, based on the level of response to adalimumab. It also shows that adalimumab was appraised side by side as part of a multi technology appraisal with etanercept. This extract raises a key question for this research: what are the factors that drive NICE coverage decisions?

This chapter provides empirical analyses of coverage decisions made by NICE during 2004-2009. First, an overview of the objectives of NICE and its appraisal process is provided. The methods for the analysis are outlined, building upon the methodological framework discussed in Chapter 3. The results of descriptive and multivariate analyses

[1] Adalimumab (Humira©) is an anti-inflammatory medicine and is indicated, among other diseases, for the treatment of adults with severe active ankylosing spondylitis (a disease causing inflammation and pain in the joints of the spine) who have not responded adequately to other treatments (European Medicines Agency, 2009, EPAR Adalimumab)

on NICE coverage decisions are then reported and explored, and limitations considered. The chapter concludes with a brief discussion about the empirical analyses performed for NICE.

### 4.1 NICE appraisal process

Established in 1999, NICE is responsible for providing guidance to the NHS in England and Wales on the funding of new technologies and their use (Chalkidou 2009). One of the key rationales for setting up NICE was to help tackle the geographic inequality in access to technology or the phenomenon more frequently referred to as 'postcode prescribing' (Summerhayes and Catchpole 2006). Since 2002 NICE's recommendations have been mandatory and NHS organizations have had to comply, usually within three months[2]. NICE "is an independent organisation responsible for providing national guidance on promoting good health and preventing and treating ill health" (NICE 2011 p.1). The NICE work programme includes not only health technology assessments of new technologies, but also of existing technologies. Its scope of work covers non-pharmaceutical technologies (e.g. medical devices) and the issuing of clinical guidelines (in addition to guidance) that provide information on how to best utilise the technology (Summerhayes and Catchpole 2006). NICE is a large organisation and combined with the network of academic assessment groups, there are over 2000 individuals contributing to NICE's guidance generation at any one time (Littlejohns et al. 2009).

Within NICE, the Centre for Health Technology Evaluation focuses on issuing guidance on the appropriate use and funding of technologies. Overall, there is a central concern for evidence-based and transparent decision-making. The NICE HTA process is characterised by involving a panel of clinical, academic, and industry stakeholders and the general public. The NICE appraisal process is governed by the use of established standard methodology for the evaluation of clinical and economic characteristics of the technology under appraisal (Littlejohns et al. 2009). The composition of the NICE Appraisal Committee includes members with health economic, statistical, clinical and academic expertise (Cairns 2006). In addition to the appraisal committee members, there are additional stakeholders that are involved in the appraisal processes, including clinical experts, patient groups, and physician/carer groups. For example, the NICE Citizens Council is a committee made up of 30 members from the general public,

[2] However, the new Conservative-Liberal Democrat government elected in May 2010 have indicated that they will change NICE's role and its coverage recommendations may only have advisory status.

aiming to provide insights into NICE decision-making through its broader public and socio-economic perspective. The technologies for appraisal by NICE - which form their work programme - are selected by the Department of Health (DH) in England and the Welsh Assembly government, although NICE is involved in specific stages of the selection and uses a series of criteria to shortlist those technologies for consideration by the DH and Welsh government (Sorenson 2008). These criteria include considerations such as the burden of disease, if the disease is a government priority, and whether there is evidence of considerable inequality in access across the UK (NICE 2008b, Sorenson 2008).

The NICE HTA process starts with a scoping exercise in which the context of the appraisal is clarified amongst the stakeholders and confirmed. This includes agreement on the technologies included in the appraisal, the comparators, the indication, and the type of clinical and economic outcomes that will be examined. Once the scoping document is confirmed, the formal appraisal process commences, and manufacturers are asked to make their submissions. Once the submissions are made, if the process followed is a Multiple Technology Appraisal (MTA), the Assessment group is involved in the reviewing of the manufacturer submission(s) and the creation of their own independent report. This information is passed on to the Appraisal Committee. If the process followed is a Single Technology Appraisal (STA), manufacturer submissions are the main source of evidence considered by the Appraisal Committee. A meeting is held to produce the Appraisal Consultation Document (ACD) which represents the draft NICE guidance. The ACD is provided to the public for consultation and comments by stakeholders, including patient groups, clinicians, and manufacturers. Following the receipt of comments on the ACD and the processing of these comments, a second Appraisal Committee meeting is held to create the Final Appraisal Determination (FAD), which represents the final guidance issued. There is also the possibility for stakeholders to appeal the FAD.

There are a range of clinical criteria evaluated during a NICE appraisal, but the main focus is on evaluating the effectiveness of the technology based on the best quality evidence. Information on the pivotal RCTs is considered of primary importance in the appraisal, and significant care is made to assess the uncertainty and generalisability of this clinical evidence. Where necessary, a formal indirect comparison using Baysian

techniques is implemented to be able to facilitate as robust as possible a comparison between the relevant comparators.

The submission of cost-utility evidence is an integral part of the NICE process, irrespective of whether an MTA or STA process is followed. This includes providing information on the cost-utility model structure, rationale, key assumptions, data inputs, main results, and sensitivity analyses surrounding the results. Detailed instructions regarding the methodology adopted for cost-utility modelling forms the basis for manufacturer submissions. Budget impact criteria are also considered – specifically to assess the incremental pharmaceutical budget implications associated with the implementation of NICE guidance.

In addition to economic and clinical criteria, the patients' perspective and patient evidence, as well as the perspective of carers and other members of the NHS are taken into consideration via the consultation and stakeholder submission process. Equity concerns are also a key factor in NICE decision making – particularly the reduction in health inequalities, and the concept of 'fairness' (NICE 2008c). To take social values into account, NICE operates through a framework that encapsulates the circumstances and the process through which social values should be incorporated in decision-making (Littlejohns et al. 2009). These eight principles include taking into account the "need to distribute health resources in the fairest way within society as a whole" (Littlejohns et al. 2009 p. 421).

## 4.2 Methods

The overall objective was to examine the impact of evidence, process and context factors on decisions made by NICE to recommend, restrict or not recommend new technologies for use in England and Wales. In addition to the general

Box 4.1 In light of the discussion in Chapter 2, the NICE-specific research hypotheses this analysis wishes to test are:

- Whether non recommendations and restrictions are increasing over time relative to recommendations.
- Whether the ICER remains a highly significant explanatory variable of NICE decision-making; specifically whether an increase in the ICER decreases the odds of recommendation.
- Whether technologies for cancer therapies will increase the odds of recommendation relative to non-recommendation.
- Whether the use of an MTA process will increase the odds of restriction relative to recommendation, compared with the STA process.
- To compare the results of NICE model with previously published models of NICE decision-making (Devlin and Parkin 2004; Dakin et al. 2006) with regards to the factors identified as having a significant impact on NICE decisions

analysis aims described in Chapter 1, the set of particular hypotheses relevant for the modeling of coverage decisions taken by NICE are highlighted in Box 4.1. Building on from the methods described in Chapter 3, this section describes the methods used to select the sample for analysis, the outcome variable and explanatory variables considered, and the statistical techniques adopted.

### *4.2.1 Sample*

The pharmaceutical technology appraisals performed by NICE formed the basis for the sample included in this analysis. The composition of the sample was determined through the following inclusion and exclusion criteria. The sample included all pharmaceutical technology appraisals (as opposed to medical devices or other interventions) conducted in the period 2004-2009 indicated for an adult population (≥18 years). To capture a sufficient number of appraisals for both individual and aggregate analyses, a five-year time horizon was implemented. Technology appraisals were excluded from the analysis for any of the following reasons: i) they focused on a non-adult population (aged <18 years); ii) the appraised technologies were non-pharmaceutical interventions; iii) marketing authorisation was withdrawn, or iv) the full guidance was not available.

### *4.2.2 Outcome variable*

To address the research question, HTA decisions were analysed by considering HTA outcomes in three categories, where the new technology can be:

- recommended for routine use
- recommended for restricted use

or

- not recommended

NICE guidance (summarised in section 1 of each guidance) indicates whether an intervention should be recommended or not for use in the NHS. A medication was considered as not recommended for use by NICE guidance if the words ‘not recommended’ were stated in section 1 of the guidance. With regards to distinguishing between recommended and restricted interventions, decision rules were developed to help classify which of the recommendations issued from NICE guidance were for routine use, and those that were for restricted use. Specifically, the Raftery (2006) classification was utilised to distinguish between restriction and recommendation.

Where a recommendation was made for a technology to be used in a population identical to its licensed indication, it was considered to be 'recommended'. Where a recommendation contained one of the following provisos in relation to the technology it was considered to be 'restricted': i) it should be used in a sub-population of its licensed indication; ii) it should be used in a second line or higher line of therapy; iii) it required monitoring; iv) it should be employed at the lowest acquisition cost; or v) it required prescription by a specialist (Raftery 2006).

### *4.2.3 Explanatory variables*

In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, the NICE dataset includes 34 variables including variables relating to (i) the clinical and economic characteristics of the technology under appraisal, as well as (ii) the processes used to come to a coverage decision, and (iii) the socio-economic context in which these decisions were made.

In addition to the common set of variables, there are certain particularities related to the appraisal process utilised by NICE. For instance, the budget impact estimates provided in NICE appraisals reflect the budgetary impact of the implementation of NICE guidance. In other words, the budget impact estimates that are reported do not always provide information on the maximum potential budget impact if the full population meeting the license criteria is eligible for treatment. To make sure that the full potential budget impact was extracted for analysis, the budget impact estimated in the Technical Appraisal Report (TAR) was used, rather than the budget impact estimated in the NICE guidance. If it was not reported in the TAR, the manufacturer's submission (where available) was used to estimate maximum potential budget impact. If that was not available, then the NICE costing template was modified to estimate what the maximum budget impact would be. Other 'NICE-specific' variables that were collected through the data extraction process included whether the appraisal was a Multiple Technology Appraisal (MTA) or a Single Technology Appraisal (STA).

### *4.2.4 Data extraction form*

The data extraction form contained the definitions and decision rules used when extracting the data from the guidance documents issued by NICE, as well as from minutes of meetings available to the public via the NICE website, and other data sources. The data extraction form was organized into three segments, relating to the

three components of analysis that are integral to this research. The data was extracted following the protocols outlined in Chapter 3, section 3.1.3. Table 4.1 provides the list of variables extracted to create the NICE dataset, as well as the accompanying decision rules, definitions and data sources.

**Table 4.1 NICE dataset: Included Variables, Definitions, Data Extraction Rules and Data Sources**

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **1** | Number of RCTs considered in decision | Count | The number of distinct Randomised Controlled Trials (RCTs) that provide data related to the therapeutic indication under evaluation.<br><br>Excluded: studies that are single arm, that have no randomization, or that are non-interventional. | Technical Appraisal Report (TAR) |
| **2** | Size of population included in RCTs | Numeric | Mean number of patients per RCT. | TAR section 2 |
| **3** | Length/extent of follow-up in RCT | Numeric | Mean number of weeks that data is collected on patients that entered the RCTs (see variable no. 1). | TAR section 2 |
| **4** | Statistically Significant results | Categorical (yes/no/inconsistent) | Presence of statistically significant superiority of technology vs. comparator for primary endpoint(s).<br><br>If more than one RCT was considered, and the technology showed statistically significant superiority in one trial, but not in another, the results were considered to be 'inconsistent' and classified as such. RCTs designed as 'non-inferiority' studies were classified as not showing any statistically significant superiority (i.e. 'no'). | TAR section 4 |
| **5** | Use of active comparator in RCTs | Numeric | Percentage of RCTs where an active comparator was used. | TAR section 4 |
| **6** | Number of observational studies considered in the guidance | Count | Number of observational studies providing information to support the study pharmaceutical. Observational studies in this circumstance are defined as studies that are non interventional (i.e. do not explicitly request the patient to take particular medication or the physician to follow a particular protocol). | TAR section 4 |
| **7** | Priority disease area | Categorical – yes/no | This variable aims to capture the health policy context in which the coverage decision is made, by capturing whether the pharmaceutical in question is linked to a disease area that is prioritized by the Department of Health. Priority disease areas were identified by examining government plans/health documents that highlight national health care system focus. | Department of Health (DH) (2002), DH (2006a), DH (2006b) DH (2007) |
| **8** | Orphan Status | Categorical – yes/no | This variable captured information on whether or not the technology was recognized by the European Medicines Agency (EMA) as an orphan designated medicine. | European Medicines Agency (accessed 2010). |
| **9** | Disease area category | Categorical – 13 categories | The British National Formulary (BNF) categories were used to classify each technology into the corresponding therapeutic area. | British National Formulary (2010) |
| **10** | Prevalence of disease/clinical | Numeric | Reported number of patients eligible for treatment, defined by patient population | TAR section 2 and/or 5; |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | condition | | for which the technology is indicated. | European Medicines Agency |
| **11** | Availability of alternative therapies in current treatment setting. | Categorical – yes/no | An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. | TAR section 2 |
| **12** | Consideration of Cost Utility Analysis in guidance | Categorical – yes/no | Presence or absence of a cost-utility analysis. | TAR section 4 |
| **13** | Incremental Cost-utility ratio of technology vs. comparator in base case | Numeric | ICER (Cost per QALY) reported in the HTA dossier for base case as accepted by the Appraisal Committee. This is defined as the ICER that is related to the recommendation.<br>If more than one ICER is presented due to the recommendation covering more than one population, then the ICER pertaining to the larger of the populations was reported.<br>In NICE appraisals, if no ICER was reported by the Appraisal Committee, then the ICER reported by the Assessment Group was used. The ICER reported in the manufacturer submission was used if it was not stated in the guidance or the assessment group report.<br>If the technology is reported as dominant or dominated, it was recorded as such in the data extraction sheet. | TAR section 4 |
| **14** | Multiple CUA/CEA models reported | Categorical - yes/no | Whether more than one cost-utility or cost-effectiveness model was considered during the appraisal | TAR section 4 |
| **15** | Multiple economic models resulting in a range of ICERs reported' | Numeric | If yes, report range of base case ICERs presented between the different models reported. The difference between the lowest and highest ICER will be calculated. | |
| **16** | Uncertainty around the base case ICER reported in submission (probabilistic) | Numeric | This should be reported as the percentage probability of acceptance at the threshold used by the agency. The probability of medication to be cost-effective at a 30,000 GBP threshold was reported. | TAR section 4/ 5 |
| **17** | Uncertainty around base case ICER reported in submission (univariate) | Numeric | This should be reported as the range of ICERs (min-max) resulting from univariate sensitivity on the base case. | TAR section 4/ 5 |
| **18** | Non-cost per QALY cost-effectiveness analyses submitted | Categorical - yes/no | Indicates if non-cost per QALY economic analyses were submitted and reviewed. | TAR section 4 |
| **19** | Anticipated budgetary impact | Numeric | Estimated annual incremental budgetary impact of introducing new medication into | TAR section 5 |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | of introduction of new technology in health care system | | the current treatment setting, if the pharmaceutical were to be introduced without any restriction. Pharmaceutical cost only (per year).<br>This is the potential budget impact were a recommendation for use in the total indicated population to be granted. | |
| **20** | Inclusion of patient submission | Categorical – yes/no | A patient submission was considered to have been included as part of the appraisal process if a submission from a patient group was posted on the webpage pertaining to the guidance. | NICE (2011) section describing the history of the appraisal |
| **21** | Number of Decision Makers Accountable | Numeric | Captures the number of decision-makers accountable for guidance issued, as reported. | TAR Appendix B of each guidance |
| **22** | Cost-effectiveness evaluation component in process | Categorical – yes/no | Captures whether cost-effectiveness is a component of the decision-making process or not. If cost-effectiveness analysis is a formal part of the appraisal process, this variable was marked as 'yes'. | TAR / NICE (2008b) |
| **23** | Budget impact as a component of decision-making process | Categorical – yes/no | Captures whether budget impact considerations are part of decision-making process | TAR / NICE (2008b) |
| **24** | Pricing known during appraisal process | Categorical – yes/no | Captures whether the price of the technology under appraisal was known during the assessment. | TAR section 3 |
| **25** | Multiple technologies or a Single technology is appraised | Categorical – MTA/STA | Data specifically extracted for NICE. Records whether guidance was issued after the appraisal of an individual technology following the STA or MTA process. As the STA process was introduced in 2006, those technologies appraised prior to this year were automatically considered as having been appraised via an MTA. | TAR Cover Page |
| **26** | Number of technologies appraised simultaneously | Count | This variable captures the number of technologies appraised simultaneously in the appraisal. | TAR cover page / section 3 |
| **27** | Accountability of pharmaceutical budget | Categorical – yes/no | The HTA agency was examined to assess whether the agency making the funding decisions is also accountable for the pharmaceutical budget or not. | Sorenson (2008); NICE (2011) |
| **28** | Independence of decision-making agency | Categorical – yes/no | This pertains to whether the HTA body is independent of the Department of Health or part of it. | Sorenson (2008); Hutton (2006); NICE (2011) |
| **29** | Date guidance was issued | Numeric | Year when coverage decision was issued | TAR cover page |
| **30** | Population size – Agency coverage | Numeric | Estimate of population size within remit of the agency performing the evaluation. | National Office of Statistics (2009) |
| **31** | GDP-healthcare expenditure | Numeric (%) | Percentage of GDP spent on healthcare, during year of decision | OECD (2009) |
| **32** | Healthcare expenditure on pharmaceuticals per patient per year | Numeric (£) | Healthcare budget spent on pharmaceuticals per patient per year, during the same year in which the appraisal was published. | Association of the British Pharmaceutical Industry (2010) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **33** | Pharmaceutical funding process within healthcare system – whether centralised or decentralised | Categorical – centralized, decentralized | States whether pharmaceutical funding process within the healthcare system is centralized at a national level or whether funding decisions are decentralized to the regional level | Sorenson (2008), Hutton (2006) |
| **34** | Election year at time of decision | Categorical – yes/no | This variable captures whether the coverage decision was made within an election year. An election year was defined as a year in which either national government or regional elections took place. | BBC 2005 |

### *4.2.5 Statistics*

The methods for the descriptive statistics and multivariate analyses were described in Chapter 3. Descriptive statistics were calculated for each extracted variable, stratified by outcome group (recommended, restricted or not recommended). Following a descriptive analysis of the dataset, a multinomial logit regression was modelled. The objective of this analysis was to obtain a parsimonious model that best reflected the main drivers of NICE decision-making.

## 4.3 Results

### *4.3.1 Sample characteristics*

**Table 4.2 Coverage decisions issued by NICE, 2004-2009**

| NICE guidance | Number of coverage decisions | Percentage |
|---|---|---|
| Recommended | 32 | 27% |
| Restricted | 69 | 58% |
| Not recommended | 17 | 14% |
| Total | 118 | 100% |

A total of 99 technology appraisals issued between January 2004 and June 2009 were retrieved from the NICE website. Of these, 65 were included in the analysis and are listed in Appendix A. Thirty-four (34) technology appraisals were excluded from the analysis for the following reasons: i) they focused on a non-adult population (n=5); they appraised non-pharmaceutical interventions (n=23); iii) marketing authorisation was withdrawn (n=2); or iv) full guidance was not available (n=4). The 65 technology appraisals that were included in the analysis covered a total of 118 technologies. Table 4.2 shows NICE guidance issued between 2004- June 2009, and the proportion of recommendations, restrictions and non recommendations made. The majority of NICE coverage decisions restricted funding for the appraised technology (58% of technologies were restricted), while the least common coverage decision was a non recommendation (14% of technologies were not recommended by NICE for NHS funding).

In the NICE sample, incomplete observations were found to occur more frequently in those variables related to the economic characteristics of the technology. The distribution of incomplete observations was concentrated in a small selection of variables linked to the economic profile of the technology. Based on this, dummy variables were inserted into the model to analyse the impact of incomplete data. A more detailed description of missing data within the NICE sample is provided in Appendix A.

### *4.3.2 Descriptive statistics*

Clinical Evidence

Descriptive statistics for the NICE sample are summarized in Table 4.3. Six variables related to the clinical evidence supporting the technology under evaluation by NICE were evaluated by decision outcome. The first three variables described the nature of the randomised clinical trial data available in terms of the number of trials, sample size and trial duration. Interventions recommended for use or for restricted use had a similar number of randomised controlled trials (RCTs) reported as part of the appraisal process (mean of 7 and 8 trials, respectively). Interventions not recommended for use had a lower number of RCTs considered in the appraisal process (mean 3 trials). The differences observed between means were statistically significant. The mean size of the patient sample included in RCTs was higher for those interventions recommended by NICE (mean = 1765 patients), compared to those interventions restricted or not recommended by NICE (1044 and 1154 patients, respectively, $p<0.05$). The mean trial duration across the three outcome groups was 96, 66 and 82 weeks respectively for recommended, restricted and not recommended interventions. The difference in mean duration between the recommended and restricted interventions was statistically significant.

Aside from the size, number and duration of RCTs, the results of the RCTs were also captured. In particular, information was collected on whether the RCTs demonstrated the technology to be statistically significantly superior in its primary endpoint relative to the comparator. 59% of recommended interventions demonstrated statistically significant superiority, as opposed to approximately 29% of restricted and 35% of not recommended interventions ($p<0.05$). The comparator used within the clinical trial programme was assessed – in particular, the percentage of comparisons made to 'active' comparators as opposed to placebo was recorded. The interventions recommended were more likely to compare to active comparators (63%) than interventions that were restricted or not recommended (40%, 44% of RCTs with active comparators, respectively. $p=0.05$).

Consideration of non-randomised observational data was also recorded. Overall, very little observational data was referred to in NICE appraisals (the mean across appraisals was 0.6 observational studies).

Disease characteristics

The prevalence of the clinical condition related to the technology under appraisal averaged at 2.2 million patients across all decisions. This ranged from 0.392 million for recommended interventions to 3.194 and 2.270 million in the restricted and not recommended interventions. The differences between outcome groups were statistically significant. The availability of alternative therapies was assessed to ascertain if it differed between recommended and restricted or not recommended interventions. In the majority of technologies appraised, an alternative was available (in 89% of cases, all decisions considered) (p=NS). It was hypothesised that there may be diseases for which technologies were more likely to be recommended rather than restricted, such as cancer therapies. In this particular sample, across the majority of disease areas there were no statistically significant differences between outcome groups. Technologies for the management of central nervous system disorders represent the exception to this rule: 22% of restricted technologies were indicated for central nervous system disorders as opposed to 6% of the recommended and not recommended technologies ($p<0.10$).

**Table 4.3 NICE Coverage Decisions 2004-2009: Descriptive statistics for extracted variables, by coverage decision (recommended, restricted, not recommended)**

| | NICE Total (n=118) | | | Recommended (n=32) | | | Restricted (n=69) | | | Not Recommended (n=17) | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | | |
| Number of RCTs considered in decision | 6.7 | 5.2 | 8.3 | 6.8 | 2.1 | 11.4 | 7.6 | 5.9 | 9.2 | 3.4 | 2.1 | 4.8 | <0.05 | 3 |
| Size of population included in RCTs | 1249 | 807 | 1691 | 1765 | 804 | 2727 | 1044 | 442 | 1645 | 1154 | 464 | 1844 | <0.05 | 3 |
| Length/extent of follow-up in RCT | 76.2 | 63.5 | 88.9 | 95.9 | 65.9 | 125.8 | 65.8 | 51.5 | 80.2 | 81.9 | 41.5 | 122.3 | NS | 1 |
| Statistically Significant results - yes | 38% | 29% | 47% | 59% | 41% | 77% | 29% | 18% | 40% | 35% | 10% | 61% | <0.05 | 2 |
| no | 15% | 9% | 22% | 6% | -3% | 15% | 17% | 8% | 27% | 24% | 1% | 46% | NS | 2 |
| inconsistent | 43% | 34% | 52% | 31% | 14% | 48% | 51% | 39% | 63% | 35% | 10% | 61% | NS | 2 |
| Relevance of RCT to payor decision | 47% | 39% | 55% | 63% | 46% | 80% | 40% | 30% | 50% | 44% | 18% | 69% | <0.10 | 3 |
| Number of observational studies considered in guidance | 0.6 | 0.1 | 1.1 | 1.5 | -0.4 | 3.4 | 0.3 | 0.1 | 0.5 | 0.0 | 0.0 | 0.0 | NS | 1 |
| Consideration of Cost Utility Analysis in guidance | 95% | 91% | 99% | 100% | 100% | 100% | 94% | 89% | 100% | 88% | 71% | 105% | NS | 2 |
| Incremental Cost-effectiveness | £31,910 | £20,945 | £42,874 | £17,782 | £11,066 | £24,498 | £24,867 | £21,002 | £28,731 | £99,239 | £11,882 | £186,597 | <0.01 | 3 |

| | NICE Total (n=118) | | | Recommended (n=32) | | | Restricted (n=69) | | | Not Recommended (n=17) | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | | |
| ratio of technology vs. comparator in base case | | | | | | | | | | | | | | |
| More than one CUA submitted | 65% | 56% | 74% | 69% | 52% | 86% | 63% | 51% | 75% | 67% | 40% | 94% | NS | 2 |
| If more than one CUA submitted - low range | £13,318 | £10,511 | £16,125 | £8,607 | £4,551 | £12,664 | £13,417 | £10,119 | £16,714 | £21,441 | £9,663 | £33,220 | <0.01 | 3 |
| If more than one CUA submitted - high range | £107,421 | £66,886 | £147,956 | £83,666 | -£12,619 | £179,951 | £129,690 | £85,202 | £174,179 | £83,242 | £33,511 | £132,974 | <0.01 | 3 |
| Uncertainty around base case ICER reported in submission (univariate) Low | £25,417 | £6,412 | £44,422 | £7,881 | £5,234 | £10,527 | £19,747 | £4,460 | £35,035 | £92,379 | -£58,911 | £243,668 | <0.01 | 1 |
| Uncertainty around base case ICER reported in submission (univariate) High | £167,389 | £56,865 | £277,913 | £113,286 | £1,293 | £225,279 | £57,146 | £29,962 | £84,329 | £731,151 | -£83,313 | £1,545,615 | <0.05 | 3 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | 43% | 34% | 52% | 61% | 45% | 77% | 41% | 29% | 53% | 8% | -1% | 17% | <0.01 | 3 |
| Non-CUA analyses | 22% | 14% | 30% | 22% | 7% | 37% | 28% | 17% | 38% | 0% | 0% | 0% | <0.10 | 2 |

| | NICE Total (n=118) | | | Recommended (n=32) | | | Restricted (n=69) | | | Not Recommended (n=17) | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | | |
| submitted | | | | | | | | | | | | | | |
| Anticipated budgetary impact of introduction of new technology in health care system | £708 | £182 | £1,235 | £36 | £18 | £54 | £828 | £51 | £1,606 | £1,632 | -£610 | £3,874 | <0.10 | 3 |
| Prevalence of disease/clinical condition | 2,288,409 | 1,189,422 | 3,387,396 | 392,063 | 82,017 | 702,109 | 3,194,243 | 1,419,388 | 4,969,097 | 2,270,955 | -290,109 | 4,832,019 | <0.01 | 3 |
| Societal Perspective adopted | 3% | 0% | 5% | 3% | -3% | 9% | 3% | -1% | 7% | 0% | 0% | 0% | NS | 2 |
| Availability of alternative therapies in current treatment setting. | 89% | 83% | 95% | 88% | 75% | 100% | 90% | 83% | 97% | 88% | 71% | 105% | NS | 2 |
| Inclusion of patient submission | 87% | 81% | 93% | 91% | 80% | 101% | 84% | 75% | 93% | 94% | 82% | 107% | NS | 2 |
| Number of Decision Makers Accountable | 30 | 28 | 32 | 28 | 25 | 31 | 31 | 29 | 34 | 28 | 23 | 33 | NS | 1 |
| Cost-effectiveness evaluation component in process | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | n/a | n/a |
| Budget impact as | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | n/a | n/a |

| | NICE Total (n=118) | | | Recommended (n=32) | | | Restricted (n=69) | | | Not Recommended (n=17) | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | | |
| a component of decision-making process | | | | | | | | | | | | | | |
| Price of technology known during appraisal | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | n/a |
| Use of STA process | 25% | 17% | 32% | 34% | 17% | 52% | 16% | 7% | 25% | 41% | 15% | 67% | <0.05 | 2 |
| Number of drugs appraised in same appraisal | 2.8 | 2.5 | 3.2 | 2.0 | 1.5 | 2.5 | 3.4 | 2.9 | 3.9 | 2.1 | 1.5 | 2.7 | <0.01 | 3 |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | n/a |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | n/a | n/a |
| Date guidance was issued | 2007 | 2006 | 2007 | 2006 | 2006 | 2007 | 2007 | 2006 | 2007 | 2007 | 2007 | 2008 | <0.05 | 1 |
| Population size – Agency coverage | 53.90 | 53.80 | 54.00 | 53.80 | 53.70 | 54.00 | 53.90 | 53.80 | 54.10 | 54.20 | 54.00 | 54.40 | <0.05 | 1 |
| GDP-healthcare expenditure | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | <0.05 | 3 |
| Healthcare expenditure on pharmaceuticals | £175 | £173 | £176 | £173 | £171 | £175 | £174 | £173 | £176 | £178 | £175 | £180 | <0.05 | 1 |
| Election year at time of decision | 7% | 2% | 11% | 6% | -3% | 15% | 7% | 1% | 14% | 6% | -7% | 18% | NS | 2 |
| Priority disease area | 56% | 47% | 65% | 59% | 41% | 77% | 57% | 45% | 69% | 47% | 21% | 74% | NS | 2 |
| Orphan Designated | 3% | 0% | 5% | 3% | -3% | 9% | 3% | -1% | 7% | 0% | 0% | 0% | NS | 2 |

| | NICE Total (n=118) | | | Recommended (n=32) | | | Restricted (n=69) | | | Not Recommended (n=17) | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | | |
| Cardiovascular system | 10% | 5% | 16% | 13% | 0% | 25% | 12% | 4% | 19% | 0% | 0% | 0% | NS | 2 |
| Central nervous system | 15% | 9% | 22% | 6% | -3% | 15% | 22% | 12% | 32% | 6% | -7% | 18% | <0.10 | 2 |
| Ear, nose and oropharynx | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | NS | 2 |
| Endocrine system | 1% | -1% | 3% | 0% | 0% | 0% | 1% | -1% | 4% | 0% | 0% | 0% | NS | 2 |
| eye | 2% | -1% | 4% | 3% | -3% | 9% | 0% | 0% | 0% | 6% | -7% | 18% | NS | 2 |
| Gastro-intestinal system | 2% | -1% | 4% | 0% | 0% | 0% | 1% | -1% | 4% | 6% | -7% | 18% | NS | 2 |
| infections | 12% | 6% | 18% | 19% | 4% | 33% | 7% | 1% | 14% | 18% | -3% | 38% | NS | 2 |
| Malignant disease and immunosuppression | 31% | 22% | 39% | 41% | 23% | 59% | 23% | 13% | 33% | 41% | 15% | 67% | NS | 2 |
| Musculoskeletal and joint diseases | 19% | 12% | 27% | 16% | 2% | 29% | 22% | 12% | 32% | 18% | -3% | 38% | NS | 2 |
| Nutrition and blood | 3% | 0% | 5% | 0% | 0% | 0% | 4% | -1% | 9% | 0% | 0% | 0% | NS | 2 |
| Obstetrics, gynaecology, and urinary-tract disorders | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | NS | 2 |
| Respiratory system | 1% | -1% | 3% | 0% | 0% | 0% | 1% | -1% | 4% | 0% | 0% | 0% | NS | 2 |
| Skin | 5% | 1% | 9% | 3% | -3% | 9% | 6% | 0% | 11% | 6% | -7% | 18% | NS | 2 |

Notes: Test 1. Both ANOVA and Kruskal-Wallis test indicate similar level of statistical significance.
Test 2. Chi-squared test used, as categorical variable.
Test 3 Either ANOVA or Kruskal-Wallis test indicate statistical significance.
NS = not significant. / N/A = not applicable

Economic Evidence

In addition to the clinical variables, a range of economic-related variables were included for analysis. The majority of NICE decisions were backed by use of the cost utility analyses (CUA) (95%). In 65% of decisions, multiple economic models were considered in the appraisal process and the use of multiple models was consistent across outcome categories. For the interventions supported by a CUA, the incremental cost-effectiveness ratio (ICER) was significantly different ($p<0.001$) between the recommended interventions (mean ICER of £17,782), compared to the restricted interventions (mean ICER of £24,498), and the interventions not recommended for use (mean ICER of £99,239). The pattern of base-case ICERs issued by NICE along with the corresponding coverage decision is presented in Figure 4.1. In general, technologies that were recommended or restricted (green and orange bars) were found in the lower ICER ranges, while not recommended technologies (red bars) were located in the upper ICER range. However, the figure also shows that, while the ICER is an important variable in NICE decision-making, there are exceptions to its usefulness as a predictor of coverage decisions. There are recommended technologies with very high ICERs, as there are technologies with lower ICERs that are not recommended. This suggests the need to examine the role of a combination of factors to explain NICE decision-making. For the interventions supported by multiple economic models, the range of ICERs reported by these models was recorded, namely the lowest and highest base-case ICERs reported across the models considered by the appraisal committee. The base-case ICERs across different models varied significantly between recommended, restricted and non-recommended interventions.

Considerable effort was made to capture information on the uncertainty around base-case ICER estimates. This was done by recording the results of probabilistic sensitivity analyses (the probability of ICER remaining below a set threshold), and by recording univariate sensitivity analyses (lowest ICER, highest ICER). The probability of the ICER remaining below £30,000 was 61% for recommended interventions, 41% for restricted interventions and 8% for interventions not recommended for use. These differences were statistically significant. With regard to univariate sensitivity analyses, the range of uncertainty was smallest for those interventions that were restricted for use (£19,747 –£ 57,146), the widest range of uncertainty was observed for interventions not recommended for use (£92,379-£731,151). These differences were statistically significant.

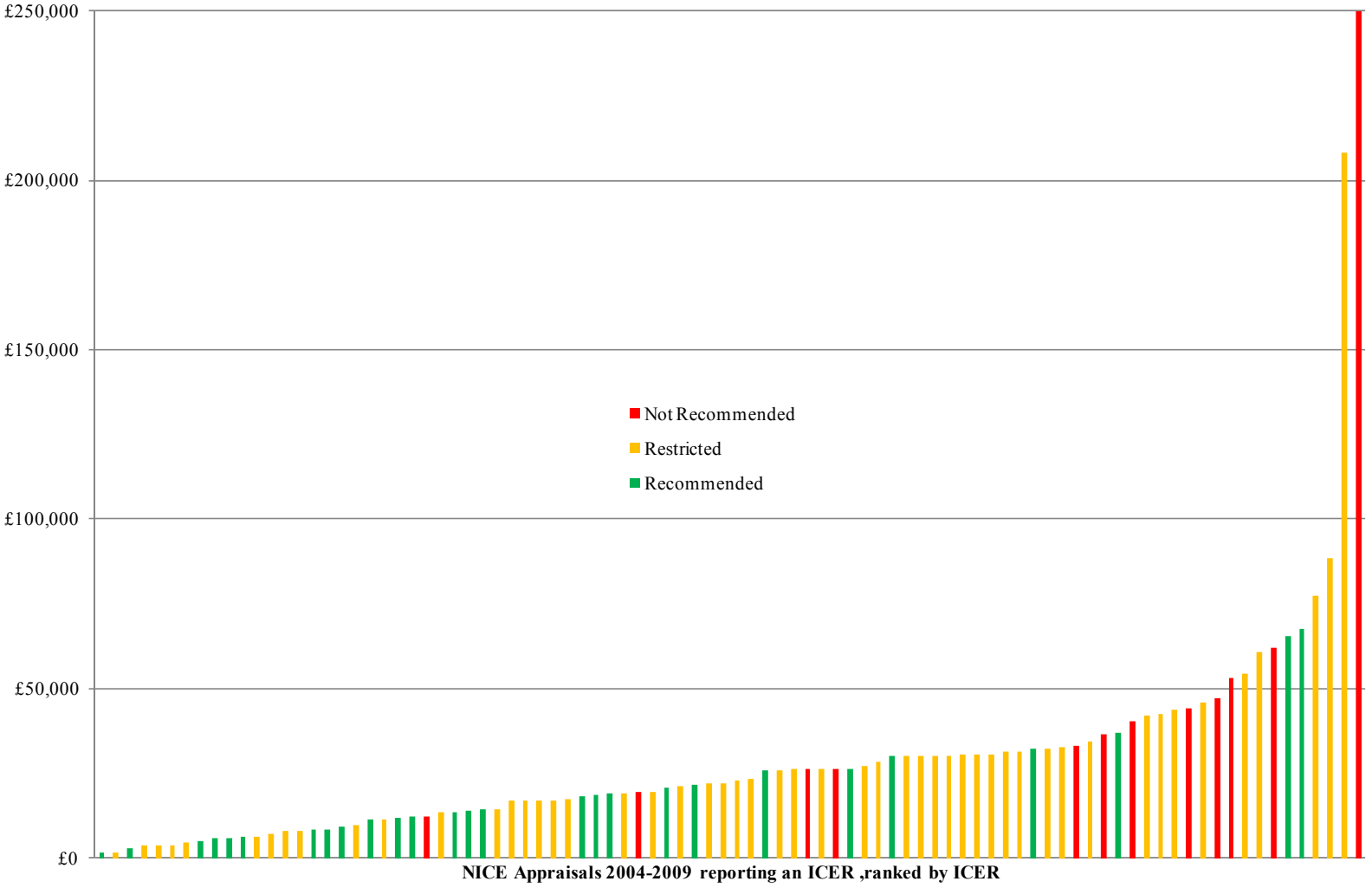
Aside from examining the results of the CUA considered by the appraisal committee, information was also captured on whether alternative cost-effectiveness models using outcomes other than cost per QALY (cost per life year gained) were considered in the decision-making process. The results varied: 22% of recommended and 28% of restricted interventions were supported by non-CUA models; however non recommended technologies were not supported by non-CUA models.

The potential budget impact of a positive recommendation was analysed - the mean estimated maximum yearly incremental budget impact across all decisions was in the order of £708 million. This ranged from £36 million for the recommended interventions to £828 and £1606 million for the restricted and not recommended interventions, respectively. While the T-test suggested a statistically significant difference between the recommended and restricted interventions ($p<0.10$), the rest of the statistical tests suggested the differences were not statistically significant.

NICE assessment process characteristics

A series of variables related to the NICE appraisal process were recorded. A societal perspective was formally referred to in the appraisal process in only 3% of decisions. In 87% of decisions, patient-group submissions were formally considered in the appraisal. On average, the appraisal committee comprised of 30 members. In all of the decisions there was a cost-effectiveness evaluation component as well as a budget impact component, and the pricing of technologies was known during the NICE appraisal process. None of these variables differed significantly between outcome groups.

**Figure 4.1 NICE: Base-case ICER reported in NICE appraisals, by coverage decision (n=90)**



Two NICE appraisal processes co-exist: the MTA process for appraising multiple technologies simultaneously and the STA process (introduced in 2006) for appraising single technologies. On average, 25% of decisions followed the STA process – although this varied across outcomes: 41% of interventions that were not recommended for NHS funding followed the STA process, as opposed to 16% of restricted interventions and 34% of recommended interventions. This difference was statistically significant ($p=0.031$). The number of technologies reviewed per appraisal was also statistically significantly different between outcome groups: restricted interventions were simultaneously appraised together with an average of 3.4 technologies compared to recommended and not recommended interventions (2 technologies appraised simultaneously on average). Statistical tests suggested the differences were statistically significant.

Socio-economic context of NICE decision-making

To complete the description of NICE decision-making in 2004-2009, the socio-economic context in which this decision-making took place was assessed. In particular, information was recorded on the year of the appraisal, size of the population under NICE's remit (approximately 53.9 million), the percentage of GDP spent on healthcare (8% on average across groups), whether appraisal coincided with a general election year and whether the technology under appraisal was linked to a disease considered a 'priority' by the Department of Health (56% of technologies appraised were directly linked to a priority disease area). There was a strong correlation between the year of appraisal and the other socio-economic factors. Across the outcome groups, the year of appraisal, population size, health care expenditure as percentage of GDP, and average health care expenditure per patient per year varied significantly.

Summary of descriptive analysis

In summary, the descriptive analysis of the NICE data set suggests that of the range of variables examined, a subset of these were found to have statistically significant differences between recommended, restricted and not recommended technologies. Of the 34 explanatory variables examined, descriptive analysis suggested that a subset of 19 variables may play an important contributing role in determining NICE decision-making. These are listed in Table 4.4.

**Table 4.4 NICE descriptive statistics: statistically significant variables (p≤ 0.05)**

| Evidence factors | Variables |
|---|---|
| Clinical Package | No. of RCTs<br>Size of RCT population<br>Duration of RCT<br>**Statistically Significant Superiority demonstrated in RCT**<br>Active comparator used<br>Consideration of observational data |
| Economic Package | **ICER**<br>Range of ICERs<br>**Uncertainty around ICER: probabilistic**, univariate<br>Use of non-CUA economic models<br>Maximum budget impact |
| Disease characteristics | Prevalence of disease<br>Technologies for treatment of CNS disorders |
| **Process factors** | **Use of STA vs. MTA process**<br>**No. of technologies reviewed simultaneously** |
| **Socio-economic context factors** | Date of appraisal<br>National population size<br>Percent GDP spent on healthcare<br>Pharmaceutical expenditure per patient per annum (GBP) |

Note: variables in bold text were statistically significant at the $p \leq 0.01$ level

### *4.3.3 Multivariate analysis results*

Following the model specification process described in Chapter 3 which included the development of a preliminary model (Appendix A), the base case NICE regression model was developed which included four variables: (i) whether statistical superiority of the primary endpoint in the RCT was demonstrated, (ii) the ICER, (iii) the number of pharmaceuticals appraised, and (iv) the year of the appraisal (Table 4.5). When the coverage decisions were regressed with these four variables, the resulting pseudo R squared was 0.26, suggesting that the four variables explained approximately 26% of the variability in NICE coverage decisions. The model suggests that a combination of clinical, economic and process variables best explain NICE decision-making. Demonstration by the technology under appraisal of statistically significant superior efficacy decreased the log odds of a restriction or non recommendation ($p=0.006$ and $p=0.016$). A unit increase in the ICER increased the log odds of moving from a recommended to a restricted decision ($p= 0.009$) versus recommendation and a not recommended decision versus recommendation ($p<0.0001$). The number of technologies reviewed simultaneously within the same appraisal had a statistically significant impact on the decision between recommendation and restriction but not between recommendation and non recommendation. The presence of an additional

technology in the appraisal process increased the log odds of restriction (p=0.005) and the log odds of non recommendation (NS). The year of appraisal also impacted significantly on the coverage decision. For every additional year, this increased the odds of a restriction (vs. recommendation) (p=0.072) and the odds of a non-recommendation (vs. recommendation) (p=0.028).

**Table 4.5 Multivariate analysis of NICE coverage decisions 2004-2009: base case model results (n=118)**

| Restricted | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -1.568 | 0.006 | -2.679 | -0.457 |
| ICER | 0.000048 | 0.009000 | 0.000012 | 0.000085 |
| Number of technologies appraised simultaneously | 0.489 | 0.005 | 0.144 | 0.834 |
| Year of Appraisal | 0.358 | 0.072 | -0.032 | 0.748 |
| Constant | -718.973 | 0.071 | -1500.761 | 62.815 |
| Not Recommended | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -2.01 | 0.02 | -3.64 | -0.38 |
| ICER | 0.000087 | <0.001 | 0.000042 | 0.000131 |
| Number of technologies appraised simultaneously | 0.12 | 0.65 | -0.39 | 0.63 |
| Year of Appraisal | 0.67 | 0.03 | 0.07 | 1.27 |
| Constant | -1351.14 | 0.03 | -2552.90 | -149.38 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.26.

Impact of alternative model specifications – sensitivity analyses

A series of sensitivity analyses were performed on the base case NICE regression model to help evaluate the robustness of the results, as outlined in Chapter 3. The general aim of these sensitivity analyses was to assess to what degree the base case model results changed if model specifications for particular assumptions were altered. The sensitivity analyses included: i) examining the impact of a binary rather than a three-category outcome variable; ii) restricting the base case analysis to complete observations, thus excluding observations with imputed values; and iii) examining the impact of assuming that the outcome variable is ordinal.

The first sensitivity analysis assessed whether the impact of the explanatory variables on NICE decision-making varied if a binary outcome variable was utilised instead of the base case three-category outcome variable (Table 4.6). In order to create this binary variable, the recommended and restricted technologies were collapsed into one category (covered), and the not recommended category was used as in the base case (not covered). The logistic regression results using a binary outcome variable no longer showed a predictive value for demonstration of clinical superiority (p=0.657).

However, in other aspects the results are similar to the base case analysis in that they confirm the impact of the ICER, number of pharmaceuticals appraised and the year of appraisal on NICE coverage decisions. Specifically, a unit increase in the ICER increased the likelihood of no coverage (p=0.005). A unit increase in the number of pharmaceuticals appraised within the same appraisal increased the odds of coverage (p=0.063). An increase in the year of appraisal appeared to increase the odds of no coverage (p=0.046).

**Table 4.6 Sensitivity Analysis 1. Multivariate analysis of NICE coverage decisions 2004-2009: alternative model using a binary outcome variable (n=118)**

| | Log Odds of Listing | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | 0.33 | 0.66 | -1.13 | 1.79 |
| ICER | -0.000043 | 0.005 | -0.000074 | -0.000013 |
| Number of technologies appraised simultaneously | 0.46 | 0.06 | -0.03 | 0.94 |
| Year of Appraisal | -0.53 | 0.05 | -1.04 | -0.01 |
| Constant | 1056.81 | 0.05 | 20.65 | 2092.97 |

In the second sensitivity analysis (Table 4.7), the regression analysis was run for a subset of complete observations (n=98 /118). This sensitivity analysis was implemented with the knowledge that removing incomplete observations from the analysis could bias the analysis. The pseudo R-squared for this model was 0.32, suggesting that this set of variables explain approximately 32% of the variability observed in NICE decision-making, as opposed to 26% in the base case model. In this sensitivity analysis, the impact of the ICER and the year of appraisal remained similar to that observed in the base case analysis, a unit increase in both parameters leading to a statistically significant increase in the log odds of a restriction or non-recommendation. However, unlike the base case model, the impact of the demonstration of statistical superiority as well as the number of technologies appraised simultaneously was weaker in this sensitivity analysis as a statistically significant effect on the odds of non-recommendation was no longer observed.

In the third sensitivity analysis, ordinality of the outcome variable was assumed. This is in contrast with the base case analysis, where ordinality was not assumed and multinomial logistic regression was used. In this sensitivity analysis ordinal logistic regression was used. The detailed results of this analysis are provided in Appendix G, and show very similar results to the base case analyses run using a multinomial logistic regression model.

**Table 4.7 Sensitivity Analysis 2. Multivariate analysis of NICE coveage decisions 2004-2009: alternative model in which incomplete observations are excluded (n=98)**

| Restricted | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -1.51 | 0.01 | -2.67 | -0.35 |
| ICER | 0.000045 | 0.02 | 0.000007 | 0.000084 |
| Number of technologies appraised simultaneously | 0.53 | 0.01 | 0.16 | 0.90 |
| Year of Appraisal | 0.51 | 0.02 | 0.07 | 0.95 |
| Constant | -1030.32 | 0.02 | -1909.69 | -150.95 |
| Not Recommended | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -1.59 | 0.12 | -3.60 | 0.42 |
| ICER | 0.00011 | <0.001 | 0.000050 | 0.00017 |
| Number of technologies appraised simultaneously | 0.08 | 0.83 | -0.65 | 0.82 |
| Year of Appraisal | 1.19 | 0.03 | 0.11 | 2.28 |
| Constant | -2397.84 | 0.03 | -4573.51 | -222.17 |

## 4.4 Discussion

The overall objective of this chapter was to examine the factors that influence decisions made by NICE to recommend, restrict or not recommend pharmaceutical technologies for use in the National Health Services of England and Wales. In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, and in light of evidence review presented in Chapter 2, a wide range of explanatory variables were included in the analysis, reflecting clinical and economic characteristics of the technology under appraisal, the appraisal process itself and the socio-economic context in which NICE operates. In addition to the general aims of the research, specific hypotheses relevant for NICE decision-making were explored and are discussed below.

The relationship between explanatory variables and the outcome variable (coverage decision) was assessed and suggested that the variation in NICE coverage decisions can be explained by four variables: whether statistical superiority of the primary endpoint in the RCT was demonstrated, the ICER, the number of pharmaceuticals appraised within the same appraisal, and the year of appraisal. The internal validity of the results obtained in this analysis was examined in two ways. Firstly, by comparing the results with published analyses of NICE decision-making, and secondly by sharing the base-case model results for review with a member of NICE (Interview with Professor Peter

Littlejohns)[1]. The aim of this interaction was to ascertain if the NICE characteristics were accurately captured in the sample used for analysis, if the approach to the analysis was clear and in particular, the reaction to the model results and potential for suggestions or additional analyses.

### *4.4.1 Pattern of NICE decision-making*

The resulting model of NICE coverage decisions provided useful glimpses into the factors driving its decision-making. The analysis of NICE coverage decisions involved the review of 118 technology appraisals performed during 2004-2009. The majority of NICE coverage decisions restricted funding to the appraised technology (58% of technologies were restricted), while the least common coverage decision was non recommendation (14% of technologies were not recommended by NICE for NHS funding). This pattern is not dissimilar to that reported in Kanavos et al. (2010) which examined NICE coverage decisions in 2007-2009: that analysis revealed that of the technologies appraised, 19% were recommended, 63% were restricted and 18% were not recommended. Clement et al. (2009) examined NICE coverage decisions between 2001- 2008 and reported 87% of technologies as listed (recommended or restricted), leaving 13% of technologies as not recommended. In the analysis of NICE decision-making during 2000-2003 produced by Dakin et al. (2006) it was reported that 21% of technologies were recommended for routine use, 66% for restricted use and 13% were not recommended. Devlin and Parkin (2004) reported NICE outcomes using a binary outcome variable, and during the period 2000-2002 reported that 71% of appraisals recommended use of the technology and 29% of appraisals did not recommend use of the technology. Therefore, the pattern observed within the data set used in this thesis seems similar to that observed in other reports, suggesting that the method of classification and data extraction used was robust.

### *4.4.2 Impact of clinical evidence and disease characteristics on NICE decision-making*

It was hypothesised that clinical variables, which were found to be significant explanatory variables in previous NICE analyses, would maintain their significance in this analysis. In the analysis of NICE decision-making presented here, demonstration

---

[1] Littlejohns, Peter. Professor, Clinical and Public Health Director National Institute for Health and Clinical Excellence. Interviewee: Karin Cerri. Interviewed by telephone, on February 2nd 2011. Meeting minutes are provided in Appendix E.

by the technology under appraisal of statistically significant superiority in its primary endpoint increased the odds of recommendation. Technologies recommended by NICE demonstrated statistically significant superior efficacy over the comparator in 59% of appraisals, compared with 29%-35% of technologies that were restricted or not recommended. This result can be seen to reflect the role of evidence-based medicine in coverage decisions and the fact that NICE defines the value of the compound in terms of the ability of the technology to demonstrate, with greater certainty, its incremental clinical value through superiority designed trials that provide stronger data to support a funding decision than technologies not able to provide evidence of superior efficacy. Comparing this result with those previously published in the literature (Dakin et al. 2006; Devlin and Parkin 2004), it is noteworthy that Devlin and Parkin (2004) did not measure the demonstration of clinical superiority while in Dakin et al. (2006), this variable was measured but was not found to have a statistically significant impact, although the sample size was smaller than that used within the analyses presented here, and pertained to NICE appraisals made in 2000-2003.

Other clinical variables which were tested in the present analysis included information on the characteristics of RCTs used in the appraisal (number, size, duration). Within descriptive analyses, these variables were found to differ statistically significantly between coverage decisions. However, they were not key drivers within the multivariate analyses. This is not to say that the characteristics of RCTs are not important or not considered within the NICE appraisal process, but that within the multivariate analysis the demonstration of superiority and the contribution of other factors, such as the ICER, had a more significant impact. This result is not dissimilar to that observed in an analysis of NICE decisions by Dakin et al. (2006), in which RCT size was not found to have statistically significant effects on the odds of a restriction or non recommendation relative to a recommendation within their multivariate analysis. The number of RCTs, however, did have a statistically significant impact in Dakin et al.'s (2006) model, which was not observed in the analyses presented here. This could be due to differences in the samples analysed: Dakin et al.'s analysis was smaller (n=60 vs. n=118 in this analysis) and the analysis included NICE appraisals published in a different time period (2000-2003), as opposed to this analyses which included appraisals from 2004-2009. The analyses by Dakin et al. (2006) also show that the use of systematic literature reviews within the NICE appraisal process appears to decrease the odds of non recommendation and restriction relative to a recommendation at

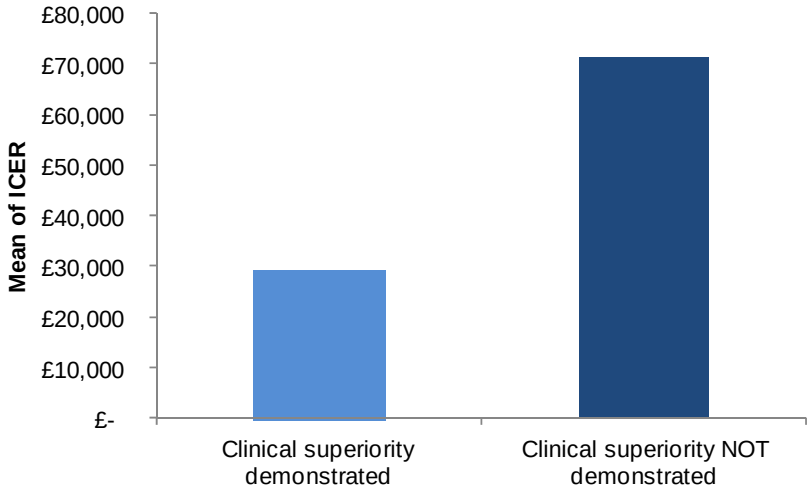
statistically significant levels. This variable was not examined in this analysis because it was noted in a review of appraisals that the majority of appraisals post 2004 included systematic reviews the clinical and economic literature as part of the process.

It was hypothesised that technologies for diseases characterised by high unmet medical need may increase the log odds of recommendation relative to non-recommendation – with particular focus on cancer therapies. Through the model specification process, the effect of cancer therapies on the log odds of non-recommendaiton was not observed and thus the final base-case model does not include this particular variable. The hypothesis was also not supported by descriptive evidence generated in this chapter. There was no statistically significant difference detected across outcome variables in the proportion of cancer therapies - of recommended techonologies, 41% were cancer therapies compared with 23% and 41% for restricted and not recommended technologies respectively.

### *4.4.3 Economic evidence and its impact on NICE decision-making*

Consistent with previously published analysis of NICE coverage decision-making (Dakin et al. 2006; Devlin and Parkin 2004), the ICER had a significant impact on NICE decision-making, and confirmed the hypothesis as stated in Box 4.1. An increase in the ICER increased the odds of a restriction versus recommendation or non recommendation versus recommendation, and this effect was highly statistically significant. The mean ICER for technologies recommended by NICE was £17,782, compared with mean ICERs of £24,867 and £99,239 for restricted technologies or technologies not recommended by NICE. The effect of the ICER was observed consistently throughout all of the sensitivity analyses conducted. This would suggest that, in addition to the strength of the clinical data, the incremental costs and benefits associated with the technology, and the resulting ICER, play a significant role in coverage decisions by NICE, and represent the agency's focus on maximization of efficiency. This interpretation of the multivariate model results is reinforced by a descriptive assessment of the mean ICER reported for technologies which demonstrated clinical superiority compared with those technologies that did not, suggesting that the ICER is a product of the clinical value demonstrated(Figure 4.2).

**Figure 4.2 NICE decision-making: mean ICER stratified by demonstration of clinical superiority in RCT**




When examining the relationship between the ICER and coverage decision in a descriptive fashion (Figure 4.1), the analysis confirms the multivariate results in two ways. Firstly, it suggests that the concept of 'value for money' is applied with some consistency in NICE decision-making such that restricted and recommended technologies tended to appear more frequently at lower ICER levels while non recommended technologies tend to appear at the upper end of the observed range of ICERs. However, the descriptive analysis also confirms the multivariate analysis by highlighting the presence of exceptions to this application of 'value for money' by NICE. There are recommended technologies with ICERs significantly above the commonly accepted threshold of £20,000 to £30,000, as there are restricted or non-recommended technologies with ICERs within commonly accepted thresholds. In part, these 'discrepancies' can be attributed to the heterogeneity of the 'restricted' category which incorporates both major and minor restrictions. This would suggest that other factors, aside from the ICER, are also driving NICE decision-making, reflecting the results of the multivariate analysis which suggests that a combination of clinical, economic, process and context factors play a role in explaining NICE decision-making.

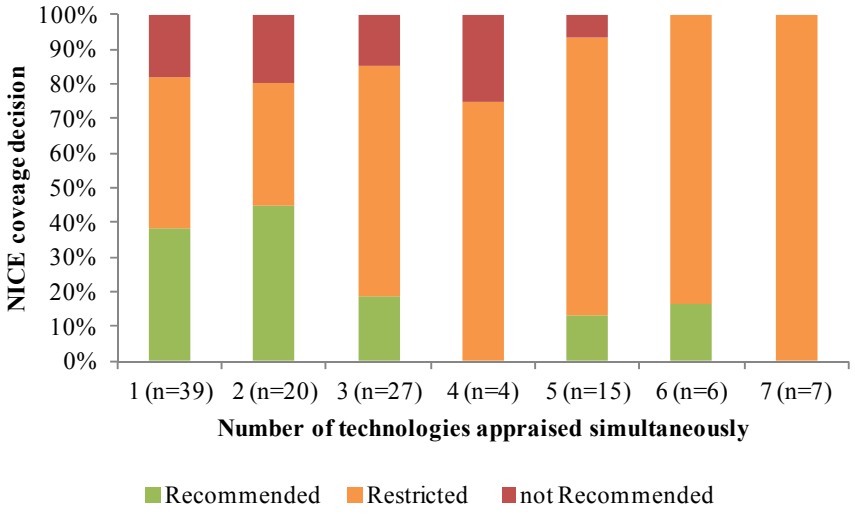
### *4.4.4 Number of technologies appraised simultaneously impacts on NICE decision-making*

The resulting multivariate model of NICE decision-making also suggests that there are process factors, beyond evidence considerations, which explain NICE coverage decisions. When considering appraisal process-related factors, the results of the model suggests that an increase in the number of technologies reviewed simultaneously within

the same appraisal increased the odds of a restriction relative to a recommendation. Technologies that were restricted were part of appraisals that on average appraised 3 technologies simultaneously, compared with recommended and not recommended technologies which were appraised with an average of 2 technologies. It was hypothesised that this may reflect the fact that NICE assessment processes differ according to the number of technologies under appraisal. Single technologies are evaluated using the STA process while multiple technologies are evaluated using the MTA process. The latter involves the use of third party assessments to provide bespoke research to support NICE assessment, while the STA process relies on manufacturer submissions similar to the process undertaken by the SMC in Scotland. However, when the model specification was altered to include a variable capturing the use of MTA or STA processes, the effect of this variable on coverage decisions was not statistically significant. Thus, the effect on coverage decisions arising from the appraisal of multiple technologies simultaneously could not be entirely explained by the use of MTA or STA alone. It was suggested by Professor P. Littlejohns[2], that the increased odds of restriction associated with higher number of technologies appraised simultaneously may reflect an approach in which a 'winner' is picked among the technologies, with the remainder recommended for restricted use or non-recommendation. To further examine Professor P. Littlejohns suggestion, a descriptive analysis was conducted of NICE decision-outcome stratified by number of technologies appraised (Figure 4.3). It suggests that that proportion of restrictions increases as the number of technologies appraised increases, and that non-recommendation appears to occur most frequently in appraisals with fewer technologies appraised simultaneously. Unfortunately, none of the published multivariate analyses of NICE decision-making examined the role of process factors, and thus a comparison with previous analyses is not possible.

[2] Littlejohns, Peter. Professor, Clinical and Public Health Director National Institute for Health and Clinical Excellence. Interviewee: Karin Cerri. Interviewed by telephone, on February 2nd 2011. Meeting minutes are provided in Appendix E.

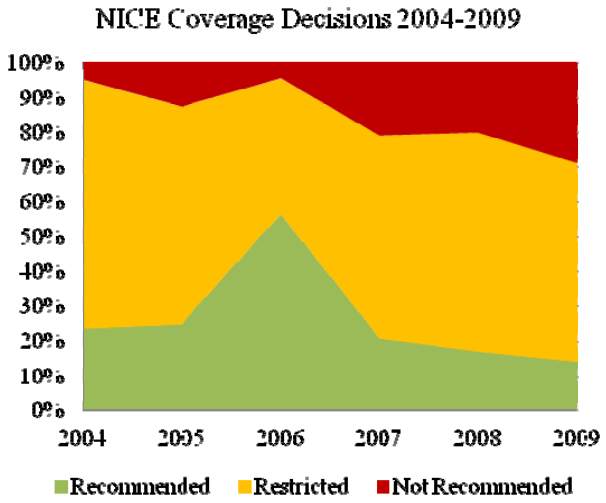
**Figure 4.3 NICE decision-outcome stratified by number of technologies appraised**




### *4.4.5 Year of appraisal impacts on NICE decision-making*

In terms of the socio-economic context of NICE decision-making, it was hypothesised that non-recommendations and restrictions are increasing over time relative to recommendations. The results of the model show that the year of appraisal impacted significantly on the coverage decision – moving from 2004 to 2009 increased the odds of a restriction versus recommendation and non-recommendation versus recommendation (Figure 4.4). The year of appraisal may reflect multiple socio-economic factors, including the political climate, a change in key staff of the HTA body, a change in societal preferences or the overall economic context. It should be noted that this analysis was conducted prior to the implementation of changes in methods of NICE appraisal process, including criteria for consideration in the appraisal of rare diseases, cancer therapies. The analysis does however cover the period where other key changes took place in NICE appraisal process including the introduction of the Single Technology Appraisal process as well as the 2008 up-date to the methods guide. By way of comparison, Dakin et al. (2006) included this variable in their analyses, and also found a statistically significant effect of the time of appraisal on outcome. In particular, an increase in the time of appraisal (i.e. as the appraisals came closer to the present), increased the odds of a non recommendation. However, Dakin et al. (2006) did not observe the effect of appraisal date on the odds of a restriction relative to a recommendation.

**Figure 4.4NICE coverage decisions between 2004-2009 (n=118)**




### *4.4.6 Limitations*

The analysis of NICE decision-making presented in this chapter was limited by several factors including: i) the dependence on publicly available information for the creation of the dataset; ii) challenges in data extraction, and iii) heterogeneity in the definition of 'restricted' technologies.

The database constructed for these analyses, incorporating information on appraisals conducted by NICE from 2004-2009, was dependent on publicly available information. Thus, it is possible that subtle concepts or rationales discussed by the appraisal committee orally were not captured in the documentation of the appraisal. In addition, the dependence on public information meant that in the situation where the information available was incomplete, it was not possible to ascertain if this was because the information was never considered in the appraisal or if it was considered but not recorded in the documentation. The presence of non-reporting reflects a lack of transparency associated with the documentation of the appraisal process. This was noted particularly with regard to the non-reporting of uncertainty information around the cost-utility/effectiveness results. Incomplete observations were taken into account in the regression models by creating dummy variables to examine whether the presence or absence of information on that variable had any explanatory value. None of these dummy variables appeared to have a significant effect on the odds of NICE coverage decisions. A sensitivity analysis was also conducted in which incomplete observations were removed from the sample for analysis. The results of this sensitivity analysis were

similar to those observed in the base case analysis, suggesting that the factors identified in the base case model are robust to alternative model specifications.

Challenges were faced in the creation of the dataset of NICE decision-making due to the dispersal of information between various documents, and the potential for inconsistency between NICE guidance drafts, manufacturer reports and Assessment Group reports, leading to a reduction in the transparency of the information that was considered by NICE. This challenge was managed by creating a specific data extraction protocol with specific rules about the nature of the data that should be extracted and how to select the data of most relevance. For example, it was not uncommon for multiple models to be submitted and considered. In these instances, the model that provided the ICER which drove the decision-making was selected and included in the dataset.

There was heterogeneity in the means through which technologies were restricted within NICE coverage decisions. The notion of restriction within NICE coverage decisions ranged from major restrictions, including restriction for use within a sub-set of the licensed indication, to minor restrictions such as the need for monitoring along with the use of the technologies. The notion of major and minor restrictions was suggested in research by Raftery (2006) on NICE guidance in which various sub-types of restrictions were presented. O'Neill and Devlin (2010) also highlighted the variation in the degree of restriction related to NICE decision-making. It is a limitation for this analysis to have such heterogeneity in the degree of restriction within a single category. However, in its actual decision-making, NICE has more than two coverage options at its disposal and thus, the use of a third coverage category within the analysis was felt to better reflect real-life HTA decision-making. To test the impact of using a binary outcome variable on the base case model results, a sensitivity analysis was performed using a binary outcome variable, which confirmed the role of the factors identified in the base case analysis.

In summary, the overall objective of this chapter was to examine the factors that influence decisions made by NICE to recommend, restrict or not recommend pharmaceutical technologies for use in its respective healthcare systems, with a focus on research hypotheses specific to NICE decision-making. The analysis provided a rich source of data from which to examine the role of each factor on NICE coverage decisions, and more importantly the contribution of each factor while adjusting for the

effect of confounding variables. The results suggest that the variability in coverage decisions observed can be explained by a combination of clinical, economic, process, and socio-economic factors. The analysis showed that the proportion of restrictions and non-recommendations issued by NICE are increasing over time relative to recommendations. The analysis also confirmed that the demonstration of clinical and economic value is central to NICE coverage decisions. While the NICE appraisal process was also shown to impact on coverage decision-making, the anticipated effect of the use of STA or MTA processes was not observed – rather an effect was observed of the number of technologies appraised simultaneously. In comparison with previously published models of NICE decision-making, there is consistency observed in the effect of clinical and economic variables, while providing new insights into the effect of variables that have not been previously studied.

## 4.5 References

Association of the British Pharmaceutical Industry (ABPI). 2010. "Facts & Statistics from the pharmaceutical industry - Medicines and the NHS" http://www.abpi.org.uk/statistics/section.asp?sect=4#15. Viewed 28 January 2010.

BBC News.2005. "2005: Historic third term for Labour" http://news.bbc.co.uk/2/hi/uk_news/politics/vote_2005/6994476.stm. Viewed 22 December 2010.

Cairns, J. 2006. Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy* 76: 134–143.

Chalkidou, K. 2009. Comparative Effectiveness Review Within the U.K.'s National Institute for Health and Clinical Excellence. *The Commonwealth Fund Issue Brief*. 59:1-12.

Clement, F. M., A. Harris, J. J. Li, K. Yong, K. M. Lee, and B. J. Manns. 2009. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 302 (13):1437-43.

Dakin, HA , NJ Devlin, and IAO Odeyemi. 2006. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. *Health Policy* 77:352-367.

Department of Health. 2002. "Improvement, expansion and reform: the next three

years priorities and planning framework 2003 – 2006" London: Crown Copyright. http://www.dh.gov.uk/dr_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4070202.pdf. Viewed 22 June 2009.

———. 2006 a. The NHS in England: the operating framework for 2006/7. Published 26 Jan 2006. London: Crown Copyright. www.dh.gov.uk/publications

———. 2006 b The NHS in England: the operating framework for 2007/08. London: Crown Copyright. www.dh.gov.uk/publications

———. 2007. The NHS in England: the Operating Framework for 2008/09 . London: Crown Copyright. www.dh.gov.uk/publications

Devlin, N., and D. Parkin. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 13 (5):437-52.

European Medicines Agency. 2009. "Adalimumab: European Public Assessment Report". http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000481/WC500050870.pdf. Viewed on 5 January 2011.

———. 2011. European public assessment reports. http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&murl=menus/medicines/medicines.jsp&mid=WC0b01ac058001d125. Viewed between 1st of June 2009-30th March 2010.

Hutton, J, C McGrath, JM Frybourg, M Tremblay, E Bramley-Harker, and C Henshall. 2006. Framework for describing and classifying decision-making systems using thechnology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal of Technology Assessment in Health Care* 22 (1):10-18.

Joint Formulary Committee. 2010. British National Formulary. 60 ed. London: British Medical Association and Royal Pharmaceutical Society. http://bnf.org/bnf/index.htm

Littlejohns, P. 2009. NICE at 10 years: new challenges ahead. *Expert Rev Pharmacoecon Outcomes Res* 9 (2):151-6.

National Institute for Health and Clinical Excellence. 2008a. NICE technology appraisal guidance 143 - Adalimumab, etanercept and infliximab for ankylosing

spondylitis. London: National Institute for Health and Clinical Excellence, 2008.

———. 2008b. Guide to the methods of technology appraisal. http://www.nice.org.uk/media/B52/A7/TA MethodsGuideUpdatedJune2008.pdf

———. 2008c. Social value judgments: principles for the development of NICE guidance. Second Edition. http://www.nice.org.uk/media/C18/30/ SVJ2PUBLICATION 2008.pdf

———. 2011. "Welcome to the National Institute for Health and Clinical Excellence". http://www.nice.org.uk/. Viewed 21 January 2011.

National Office for Statistics. 2009. "Key demographic and health indicators, 1976 onwards: Population Trends. ONS 1976-2008 data". http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=9537 . Viewed 23 June 2009.

OECD. 2009. "OECD Health Data - Version: November 09". http://www.ecosante.org/index2.php?base=OCDE&langs=ENG&langh=ENG. Viewed 22 June 2010.

O'Neill, P., and N. J. Devlin. 2010. An analysis of NICE's 'restricted' (or 'optimized') decisions. *Pharmacoeconomics* 28 (11):987-93.

Sorenson, C., Drummond, M., and Kanavos, P. 2008. Ensuring value for Money in Health Care: the role of HTA in the European Union. Cornwall: World Health Organization 2008, on behalf of the European Observatory on Health Systems and Policies.

Summerhayes, M., and P. Catchpole. 2006. Has NICE been nice to cancer? *Eur J Cancer* 42 (17):2881-6.

Raftery, J. 2006. Review of NICE's recommendations, 1999-2005. *BMJ* 332 (7552):1266-8.

# 5 Empirical analysis of SMC coverage decisions

Adalimumab is indicated for the treatment of ankylosing spondylitis[1], and was reviewed by NICE, SMC, CVZ and HAS to assess whether it should be funded by the healthcare system. The SMC provided the following guidance:

> "Adalimumab (Humira©) is accepted for restricted use within NHS Scotland for the treatment of adults with severe active ankylosing spondylitis who have an inadequate response to conventional therapy. It is restricted to use in accordance with the British Society for Rheumatology (BSR) guidelines of July 2004.
>
> Adalimumab improves signs, symptoms, physical function and quality of life in patients with severe active ankylosing spondylitis. It reduces spinal inflammation, but there is no radiological evidence that it decreases joint damage. An economic evaluation demonstrated that it is a cost-effective treatment option when used in tumour necrosis factor (TNF)-antagonist naïve patients in accordance with the BSR guidelines and where clear and rigorous stopping rules are applied."
> (SMC 2006 p. 1)

The guidance issued by the SMC on adalimumab is not dissimilar to that issued by NICE (see Chapter 4), in which its use is restricted to a sub-set of the eligible population that have failed at least two conventional therapies. While arriving at an apparently similar coverage decision to that of NICE, the SMC assessment differed from the NICE assessment in the evidence, process and context within which it made its decision. For example, the SMC considered one economic model, while NICE considered five models (NICE 2008); the SMC analysed adalimumab alone, while NICE analysed adalimumab within the context of a Multi Technology Appraisal (MTA) process. What is the impact of evidence, process and context factors on SMC decisions? Do the factors that drive NICE decisions, as highlighted in Chapter 4, play a role in SMC coverage decisions? Or is SMC decision-making best explained by a different mix of evidence, process and socio-economic context factors?

This chapter provides an empirical analysis of coverage decisions made by SMC during the period 2004-2009. First, an overview of SMC, its objectives and appraisal process

---

1 Adalimumab (Humira©) is an anti-inflammatory medicine and is indicated, among other diseases, for the treatment of adults with severe active ankylosing spondylitis (a disease causing inflammation and pain in the joints of the spine) who have not responded adequately to other treatments (European Medicines Agency, 2009, EPAR Adalimumab)

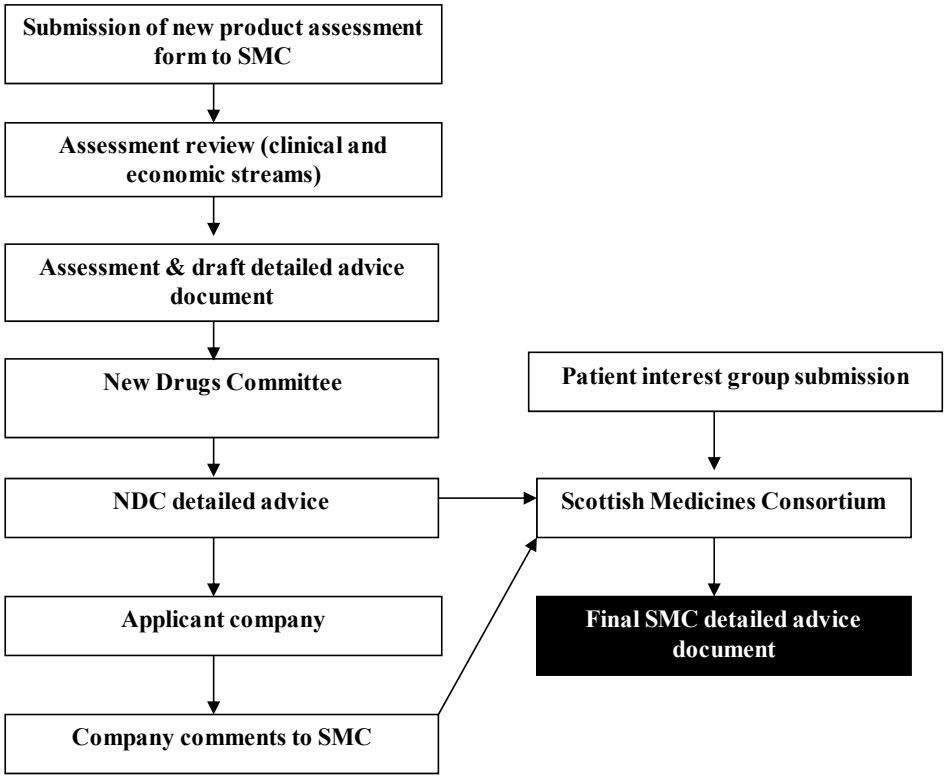
is provided. The methods for the analysis are then outlined, building upon the methods discussed in Chapter 3. Next, the results of descriptive and multivariate analyses of SMC coverage decisions are reported and explored, and limitations considered. The chapter concludes with a brief discussion on the empirical analyses performed for SMC.

## 5.1 SMC Appraisal Process

Established in 2001, the Scottish Medicines Consortium advises NHS Scotland on the use of newly licensed technologies or newly licensed indications for existing technologies. The purpose of the SMC is "to avoid duplication of new medicines assessment by individual ADTCs [NHS Board Area Drugs and Therapeutics Committees], to avoid geographical inequity in decision making and to make the best use of expertise available across Scotland" (SMC 2010a p. 4). The SMC is made up of a consortium of 14 Health Boards. The advice of the SMC is made available to the health boards, which then make a recommendation as to the use of technologies within their area of remit. Individual clinicians make the final prescribing decision in relation to their patients. In addition to the evaluation of new technologies and indications, the SMC also has a role in horizon scanning – that is, the identification of new technologies in development and the communication of such technologies to the Scottish NHS Health Boards to facilitate financial and service planning for future years.

The SMC provides guidance to NHS Scotland based on a rapid review soon after the marketing authorisation is obtained for the technology (Cairns 2006; SMC 2010a). The assessment of the SMC is based on evidence submitted by the manufacturer, and there is no third party assessment within its appraisal process (Drummond and Mason 2009). The New Drugs Committee (NDC) is a sub-group of the SMC that provides an assessment of the clinical and economic evidence and formulates a draft advice for consideration by the SMC. The SMC issues the final advice after consultation with its members and stakeholders. The composition of the Consortium and the NDC that advises the SMC on new technologies includes clinicians, pharmacists and health economists, in addition to stakeholders from manufacturers, patient groups and the government. The appraisal process and key stakeholders are highlighted in Figure 5.1.

**Figure 5.1 SMC appraisal process**




*Source*: SMC 2010a.

A series of clinical criteria are taken into account in the SMC's assessment of pharmaceutical technologies (SMC 2010b). This includes information on the evidence demonstrating the efficacy of the technology, with a particular focus on randomised controlled studies (RCTs). The importance of evidence from active-controlled studies, as opposed to placebo-controlled studies, is stressed. The process however, also allows for the submission of non-randomised studies, although the rationale for doing so needs to be provided. In addition to efficacy data, safety information is also assessed. Clinical effectiveness, relative to a comparator, is a key part of the SMC appraisal, which aims to assess the relevance of the efficacy and safety outcomes of the clinical studies to treatment practice and patients in Scotland.

A range of economic criteria are included in the SMC appraisal process. The pharmacoeconomic assessment examines the economic model and results submitted by the manufacturer, including its relevance and robustness. Sensitivity analysis around the economic model results is an important component of the evaluation. In addition to pharmacoeconomic assessments, resource implications associated with the introduction of the technology are also assessed by the SMC. The assessment is primarily focused on estimating the size of the target population and the incremental drug budget

associated with the introduction of the technology. In addition to clinical and economic criteria, submissions by patient groups are a formal part of the appraisal process. Patient groups have the option to provide information via a specific submission form, which is then considered by the SMC during its appraisal. Having outlined the role and appraisal process of the SMC, the next section summarises the methods used to analyse the factors driving SMC coverage decisions.

## 5.2 Methods

The overall objective was to examine the factors that drive decisions made by SMC to accept for use, accept for restricted or not recommend. In addition to the general analysis aims described in Chapter 1, the particular hypotheses relevant for the analysis of coverage decisions by SMC are highlighted in Box 5.1. Building on from the methods described in Chapter 3, this section describes the methods used to select the sample for analysis, the outcome variable and explanatory variables considered, and the statistical techniques adopted.

Box. 5.1 In light of the discussions presented in Chapter 2, SMC-specific research objectives were to test whether:

- The odds of SMC recommendation is impacted upon by the type of submission (first-time submission or resubmission)
- The pattern of coverage decision-making by SMC is changing over time – such that the proportion of non-recommendation and restrictions were increasing over time relative to recommendations
- The ICER is a key explanatory variable of SMC decision-making and increasing ICERs are hypothesised to be associated with an increased log odds of restriction or non-recommendation relative to recommendation
- Pharmaceutical budget impact estimates significantly impact on SMC outcomes – increasing budgetary impact is hypothesised to increase the log-odds of non-recommendation or restriction relative to recommendation
- The odds of non-recommendation decrease for technologies indicated for conditions with no alternative treatment options

### *5.2.1 Sample*

Pharmaceutical technology assessments performed by the SMC formed the basis of the sample included in this analysis. The composition of the sample was determined through the following inclusion and exclusion criteria. The sample included all drug technology appraisals (as opposed to medical devices or other interventions) completed in 2004-2009 and indicated for an adult population (≥18 years). Technology assessments were excluded from the analysis for any of the following reasons: i) they focused on a non-adult population; ii) they appraised non-drug interventions; iii) marketing authorisation was withdrawn; iv) an abbreviated or IRP guidance was issued; or v) the full advice report was not available.

### *5.2.2 Outcome variable*

To address the research question, SMC decisions were analysed by considering SMC decision outcomes in three categories:

- accepted for routine use
- accepted for restricted use
- not recommended for use

The SMC is the only one of the four HTA bodies examined in this thesis that explicitly uses these three categories to express the outcome of its decision-making and these correspond closely with the categories needed for the analyses. Thus, no additional algorithm or decision rule was used to translate coverage decisions made by the SMC into the three categories used for descriptive and multivariate analysis.

### *5.2.3 Explanatory variables*

In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, the SMC dataset includes 36 explanatory variables, including variables collecting information on i) the clinical, disease and economic characteristics of the technology under appraisal, as well as ii) the process used to come to a coverage decision, and iii) the socio-economic context in which these decisions were made.

### *5.2.4 Data extraction form*

Similar to the form utilised to extract data from NICE guidance (see Chapter 4), the SMC form contained the definitions and decision rules used when extracting the data from the guidance documents issued by SMC, as well as from minutes of meetings available to the public via the SMC website, and other data sources. The data extraction form was organized into three segments, relating to the three components of analysis that are integral to this research. The data was extracted following the protocols outlined in Chapter 3 (Section 3.1.3). Table 5.1 provides the list of variables extracted to create the SMC dataset, as well as the accompanying decision rules and definitions.

**Table 5.1 SMC dataset: Included Variables, Definitions, Data Extraction Rule and Data Sources**

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **1** | Number of RCTs considered in decision | Count | The number of distinct RCTs that provide data related to the therapeutic indication under evaluation.<br><br>Excluded: studies that are single arm, that have no randomization, or that are non-interventional. | SMC Advice, "Summary of evidence on comparative efficacy" section |
| **2** | Size of population included in RCTs | Numeric | Mean number of patients per RCT. | SMC Advice, "Summary of evidence on comparative efficacy" section |
| **3** | Length/extent of follow-up in RCT | Numeric | Mean number of weeks that data is collected on patients that entered the RCTs (see variable no. 1). | SMC Advice, "Summary of evidence on comparative efficacy" section |
| **4** | Statistically Significant results | Categorical (yes/no/inconsistent) | Presence of statistically significant superiority of technology vs. comparator for primary endpoint(s).<br><br>If more than one RCT was considered, and the technology showed statistically significant superiority in one trial, but not in another, the results were considered to be 'inconsistent' and classified as such. RCTs designed as 'non-inferiority' studies were classified as not showing any statistically significant superiority (i.e. 'no'). | SMC Advice, "Summary of evidence on comparative efficacy" section |
| **5** | Use of active comparator in RCT | Numeric | Percentage of RCTs where active comparator was used. | SMC Advice, "Summary of evidence on comparative efficacy" section and " Summary of clinical effectiveness issues" section |
| **6** | Number of observational studies considered in guidance | Count | Number of observational studies providing information to support study drug.<br><br>Observational studies in this circumstance are defined as studies that are non interventional (i.e. do not explicitly request the patient to take particular medication or the physician to follow particular protocol). | SMC Advice, "Summary of evidence on comparative efficacy" section, and " Summary of evidence on comparative safety" section |
| **7** | Priority disease area | Categorical – yes/no | This variable aims to capture the health policy context in which payer decision is made, by capturing whether the pharmaceutical in question is linked to a disease area that is prioritized by the Department of Health. Priority disease areas were identified by examining government plans/health documents that highlight national health care system focus. | NHS Health Scotland (2004, 2005, 2006, 2007, 2008) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **8** | Orphan Status | Categorical – yes/no | This variable captured information on whether or not the technology was recognized by the European Medicines Agency (EMA) as an orphan designated medicine. | European Medicines Agency (accessed 2010) |
| **9** | Therapeutic Area | Categorical – 13 categories | The British National Formulary (BNF) categories were used to classify each technology into the corresponding therapeutic area. | British National Formulary (2010) |
| **10** | Prevalence of disease/clinical condition | Numeric | Reported number of patients eligible for treatment, as per the Summary Product Characteristics and indication of the medication under evaluation. | SMC Advice, " Additional information: budget impact" section |
| **11** | Availability of alternative therapies in current treatment setting. | Categorical – yes/no | An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. | SMC Advice " Additional information: comparators" section |
| **12** | Consideration of Cost Utility Analysis in guidance | Categorical – CUA performed or no CUA | Presence or absence of a cost-utility analysis. | SMC Advice " Summary of comparative health economic evidence" section |
| **13** | Incremental Cost-utility ratio of technology vs. comparator in base case | Numeric | ICER (Cost per QALY) reported in the HTA dossier for base case as accepted by the Appraisal Committee. This is defined as the ICER that is related to the recommendation.<br><br>If more than one ICER is presented as the recommendation covers more than one population, than the ICER pertaining to the larger of the populations was reported.<br><br>If technology is reported as dominant or dominated, it was recorded as such in data extraction sheet. | SMC Advice " Summary of comparative health economic evidence" section |
| **14** | Multiple CUA/CEA models reported | Categorical - Yes/No | Whether more than one cost-utility or cost-effectiveness model was considered during the appraisal | SMC Advice " Summary of comparative health economic evidence" section |
| **15** | Multiple economic models resulting in a range of ICERs reported | Numeric | If yes, report range of base case ICERs presented between the different models reported. The difference between the lowest and highest ICER will be calculated. | SMC Advice " Summary of comparative health economic evidence" section |
| **16** | Uncertainty around the base case ICER reported in submission (probabilistic) | Numeric | This should be reported as the percentage probability of acceptance at the threshold used by the agency. The probability of medication to be cost-effective at a 30,000 GBP threshold was reported. | SMC Advice " Summary of comparative health economic evidence" section |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **17** | Uncertainty around base case ICER reported in submission (univariate) | Numeric | This should be reported as the range of ICERs (min-max) resulting from univariate sensitivity on the base case. | SMC Advice " Summary of comparative health economic evidence" section |
| **18** | Non-cost per QALY cost-effectiveness analyses submitted | Categorical - Yes/No | Indicates if non-cost per QALY economic analyses were submitted and reviewed. | SMC Advice " Summary of comparative health economic evidence" section |
| **19** | Anticipated budgetary impact of introduction of new technology in health care system | Numeric | Estimated annual budgetary impact of introducing new medication into the current treatment setting, if the pharmaceutical were to be introduced without any restriction. Drug cost only (per year). This is the potential budget impact were a recommendation for use in the total indicated population is granted. | SMC Advice, " Additional information: budget impact" section |
| **20** | Availability of alternative therapies in current treatment setting. | Categorical – yes/no | An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. | SMC Advice " Additional information: comparators" section |
| **21** | Inclusion of patient submission | Categorical – yes/no | A patient submission was considered to have been included as part of the appraisal process if a submission from a patient group was acknowledged to have been included in the relevant section of the SMC Advice. | SMC Advice, " Summary of patient and public involvement" section |
| **22** | Number of decision makers accountable | Numeric | Captures the number of decision-makers accountable for guidance issued, as reported in meeting minutes in which the technology was discussed. | SMC Meeting Minutes, http://www.scottishmedicines.org.uk/About_SMC/Minutes/Minutes |
| **23** | Cost-effectiveness evaluation component in process | Categorical – yes/no | Captures whether cost-effectiveness is a component of the decision-making process or not. If cost-effectiveness analysis is a formal part of the appraisal process, this variable was marked as 'yes'. | SMC Advice, SMC (2010b) |
| **24** | Budget impact as a component of decision-making process | Categorical – yes/no | Captures whether budget impact considerations are part of decision-making process | SMC Advice, SMC (2010b) |
| **25** | Pricing known during appraisal process | Categorical – yes/no | Captures whether the price of the technology under appraisal was known during the assessment. | SMC Advice, SMC (2010b) |
| **26** | Number of technologies appraised simultaneously | Count | This variable captures the number of technologies appraised simultaneously in the appraisal. | SMC Advice, SMC (2010b) |
| **27** | Different process for medications destined for | Categorical – yes/no | Records whether funding decisions for medications follow different processes depending on whether they are destined for hospital or retail | SMC Advice, SMC (2010b) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | hospital or retail use | | prescription. | |
| **28** | Accountability of drug budget | Categorical – yes/no | The HTA agency was examined to assess whether the agency making the funding decisions is also accountable for the drug budget or not. | SMC (2010a) |
| **29** | Independence of decision-making agency | Categorical – yes/no | This pertains to whether the HTA body is independent of Department of Health or part of it. | SMC (2010a) |
| **30** | Date guidance was issued | Numeric | Year when coverage decision was issued | SMC Advice |
| **31** | Population size – Agency coverage | Numeric | Estimate of population size within remit of the agency performing the evaluation. | National Office of Statistics (2009) |
| **32** | GDP-healthcare expenditure | Numeric (%) | Percentage of GDP spent on healthcare, during year of decision | OECD (2009) |
| **33** | Healthcare expenditure on pharmaceuticals per patient per year | Numeric (£) | Healthcare budget spent on pharmaceuticals per patient per year, during the same year in which the appraisal was published. | ISD (2009) |
| **34** | Drug funding process within healthcare system – whether centralised or decentralised | Categorical – centralised, decentralised | States whether drug funding process within the healthcare system is centralized at a national level or whether funding decisions are decentralized to the regional level | SMC (2010a) |
| **35** | Election year at time of decision | Categorical – yes/no | This variable captures whether payer decision was made within an election year. An election year was defined as a year in which either national government or regional elections took place. | BBC 2005, BBC 2007 |
| **36** | Submission type | Categorical – first or re-submission | Collected for the SMC – this provides information on whether the submission under analysis is a first full submission or re-submission. This data was extracted directly from the appraisal summary document. | SMC Advice cover page |

### *5.2.5 Statistics*

The methods for the descriptive statistics and multivariate analyses were described in Chapter 3. Descriptive statistics were calculated for each extracted variable, stratified by outcome group (recommended, restricted or not recommended). Following a descriptive analysis of the dataset, a multinomial logit regression was modelled. The objective of this analysis was to obtain a parsimonious model that best reflected the main drivers of SMC decision-making.

## 5.3 Results

### *5.3.1 Sample characteristics*

**Table 5.2 Outcome of SMC Guidance issued between 2004-June 2009**

| SMC guidance | Number of coverage decisions | Percentage |
|---|---|---|
| Recommended | 54 | 19% |
| Restricted | 102 | 35% |
| Not recommended | 132 | 46% |
| Total | 288 | 100% |

A total of 531 drug reviews issued between January 2004 and June 2009 were retrieved from the SMC website. Of these, 288 full submissions and resubmissions were included for analysis. A total of 243 technology appraisals were excluded from the analysis for the following reasons: i) due to an abbreviated/IRP or non submission (n=169) ii) full guidance was not available (n=58); iii) they focused on a non-adult population (n=11); and iv) marketing authorisation was withdrawn (n=5). Prior to January 2005, full guidance was not reported, therefore this analysis focused on SMC coverage decisions made between 2005-2009. The 288 drug reviews that were included in the analysis covered a total of 184 medications. The most common coverage decision by the SMC was to not recommend funding for the new technology (46%), followed by restriction of funding (35% of technologies), and technologies recommended for funding (19%) (Table 5.2).

In the SMC sample (n=288), there were about 15% of variable entries for which data was lacking. The variables with the highest number of incomplete entries were for those variables related to the economic characteristics of the technology. More than half of submissions where cost-utility models were presented did not report uncertainty

estimates around the base case ICER. A more detailed description of incomplete data within the SMC sample is provided in Appendix B.

### *5.3.2 Descriptive statistics*

Clinical Evidence

Descriptive statistics for the SMC sample are summarised in Table 5.3. Six variables related to the clinical evidence supporting the technology under evaluation by SMC were evaluated by decision outcome. The first three variables described the nature of the randomised clinical trial data available in terms of the number of trials, sample size and trial duration. On average SMC took into consideration 2 RCTs in their review process. This did not differ between interventions that were recommended, restricted or not recommended. The mean size of the patient sample included in RCTs was higher for those interventions recommended by SMC (mean = 1532 patients), compared to those interventions restricted or not recommended by SMC (1360 and 484 patients, respectively) ($p<0.01$). The mean trial duration across the three outcome groups was 60, 50 and 35 weeks respectively for recommended, restricted and not recommended interventions ($p<0.05$).

The impact of the RCT design and outcomes was also assessed. In particular, information was collected as to whether the RCTs demonstrated the technology to be statistically significantly superior in its primary endpoint to the comparator. Unexpectedly, a higher proportion of RCTs demonstrating superiority of efficacy were not recommended. 48% of recommended interventions demonstrated statistically significant superiority, as opposed to approximately 52% of restricted and 58% not recommended interventions (NS). The interventions recommended for use had a higher chance of comparison to active comparators (67%) than interventions that were restricted or not recommended (53%, 43% of RCTs with active comparators, respectively) ($p<0.05$).

Consideration of non-randomised observational data was also recorded. On average, SMC appraisals considered 1.3 observational studies. Recommended and restricted interventions did not consider any observational data, compared to not recommended interventions supported by 2.7 studies. However, these differences between outcome groups were not statistically significant.

Disease characteristics

Disease characteristics, including the disease category, size of the population eligible for treatment, the availability of alternative therapies, and orphan designation status were recorded. The average size of the eligible population for treatment by the technology under review was 11,229 patients across all decisions. This ranged from 36,122 patients for recommended interventions to 6,584 and 4,000 patients in the restricted and not recommended interventions ($p<0.05$).

The availability of alternative therapies was assessed to ascertain if this differed between recommended and restricted or not recommended interventions. In the majority of technologies appraised, an alternative was available (in 83% of cases, all decisions considered). Unexpectedly, the availability of alternative therapies was higher in the recommended group (93%), than in the not recommended group (77%). It had been hypothesised that therapies indicated for treatment of diseases for which there was no alternative would be more likely to be recommended due to higher unmet need. This did not appear to be the case, based on the descriptive analyses presented. The differences between outcomes were statistically significant ($p= 0.031$). Of the technologies assessed by the SMC, 13% had an orphan designation. A lower proportion of orphan technologies were in the recommended group (6%) relative to the restricted (9%) and not recommended group (18%). These differences were statistically significant between the groups, suggesting that a higher proportion of orphan designated technologies were not recommended. A higher proportion of technologies indicated for infectious diseases were in the non recommended group (5%) compared with the recommended (17%) and restricted groups (16%) ($p<0.05$).

**Table 5.3 SMC Coverage Decisions: Descriptive statistics for extracted variables, by coverage decision (recommended, restricted, not recommended)**

| | **Total SMC sample (n=288)** | | | **SMC Recommended Technologies (n=54)** | | | **SMC Restricted Technologies (n=102** | | | **SMC Not Recommended Technologies (n=132)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | |
| Number of RCTs | 2.2 | 1.9 | 2.5 | 2.3 | 1.7 | 2.9 | 2.3 | 1.7 | 2.8 | 2.1 | 1.8 | 2.4 |
| Size of population included in RCTs | 999 | 693 | 1305 | 1532 | 572 | 2492 | 1360 | 691 | 2028 | 484 | 377 | 592 |
| Length/extent of follow-up in RCT (weeks) | 45 | 39 | 52 | 60 | 42 | 78 | 50 | 38 | 63 | 35 | 28 | 42 |
| Statistically Significant results - yes | 54% | 48% | 60% | 48% | 34% | 62% | 52% | 42% | 62% | 58% | 49% | 66% |
| no | 21% | 16% | 26% | 22% | 11% | 34% | 23% | 14% | 31% | 20% | 13% | 27% |
| inconsistent | 18% | 13% | 22% | 24% | 12% | 36% | 22% | 13% | 30% | 12% | 6% | 18% |
| Use of Active Comparator in RCT | 51% | 45% | 57% | 67% | 55% | 79% | 53% | 44% | 62% | 43% | 34% | 51% |
| Number of observational studies considered in guidance | 1.3 | -0.4 | 3.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.2 | 2.7 | -1.0 | 6.4 |
| Consideration of Cost Utility Analysis in guidance | 74% | 69% | 79% | 67% | 54% | 80% | 77% | 69% | 86% | 75% | 68% | 82% |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | £34,013 | £21,661 | £46,364 | £11,893 | £8,645 | £15,140 | £26,316 | £13,265 | £39,367 | £46,679 | £23,373 | £69,985 |
| More than one CUA submitted | 1% | 0% | 2% | 0% | 0% | 0% | 1% | -1% | 3% | 2% | -1% | 4% |
| If more than one CUA submitted - low range | £10,399 | -£9,782 | £30,580 | . | . | . | £3,194 | . | . | £14,002 | -£52,096 | £80,100 |

| | Total SMC sample (n=288) | | | SMC Recommended Technologies (n=54) | | | SMC Restricted Technologies (n=102 | | | SMC Not Recommended Technologies (n=132) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | |
| If More than one CUA submitted - high range | £18,207 | -£2,672 | £39,086 | . | . | . | £10,920 | . | . | £21,851 | -£48,681 | £92,383 |
| Uncertainty around the base case ICER (probabilistic) | 58% | 43% | 72% | 72% | -164% | 307% | 71% | 53% | 90% | 37% | 13% | 61% |
| Uncertainty around base case ICER (univariate) Low | £33,277 | £8,916 | £57,637 | £11,251 | £5,017 | £17,486 | £8,963 | £5,794 | £12,132 | £63,494 | £8,516 | £118,471 |
| Uncertainty around base case ICER (univariate) High | £77,071 | £27,230 | £126,912 | £31,647 | £19,847 | £43,448 | £27,112 | £20,849 | £33,375 | £147,174 | £26,746 | £267,601 |
| Non-CUA analyses submitted | 30% | 25% | 36% | 39% | 25% | 52% | 25% | 16% | 33% | 31% | 23% | 39% |
| Potential budgetary impact (million) | £1.190 | £0.928 | £1.451 | £1.816 | £0.923 | £2.708 | £1.205 | £0.743 | £1.668 | £0.907 | £0.661 | £1.153 |
| Prevalence of disease/clinical condition | 11229 | 1640 | 20819 | 36122 | -12168 | 84412 | 6584 | 1612 | 11557 | 4000 | 1694 | 6307 |
| Societal Perspective adopted | 0% | 0% | 1% | 0% | 0% | 0% | 1% | -1% | 3% | 0% | 0% | 0% |
| Availability of alternative therapies in current treatment setting. | 83% | 79% | 87% | 93% | 85% | 100% | 85% | 78% | 92% | 77% | 70% | 85% |
| Inclusion of patient submission | 41% | 35% | 48% | 27% | 14% | 41% | 45% | 34% | 56% | 44% | 35% | 54% |
| Number of Decision Makers Accountable | 25 | 24 | 25 | 25 | 24 | 26 | 25 | 24 | 25 | 25 | 24 | 26 |
| Cost-effectiveness evaluation component in | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

| | **Total SMC sample (n=288)** | | | **SMC Recommended Technologies (n=54)** | | | **SMC Restricted Technologies (n=102** | | | **SMC Not Recommended Technologies (n=132)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | |
| process | | | | | | | | | | | | |
| Budget impact as a component of decision-making process | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Price of technology known during appraisal | 100% | 100% | 100% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Date guidance was issued | 2007 | 2007 | 2007 | 2007 | 2006 | 2007 | 2007 | 2006 | 2007 | 2007 | 2007 | 2007 |
| Population size – Agency coverage (millions) | 5137 | 5133 | 5140 | 5138 | 5129 | 5147 | 5135 | 5129 | 5141 | 5138 | 5133 | 5143 |
| GDP-healthcare expenditure | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% | 8% |
| Healthcare expenditure on pharmaceuticals | £191 | £191 | £192 | £191 | £190 | £193 | £191 | £190 | £192 | £192 | £191 | £193 |
| Election year at time of decision | 47% | 41% | 53% | 48% | 34% | 62% | 48% | 38% | 58% | 45% | 37% | 54% |
| Priority disease area | 66% | 61% | 72% | 67% | 54% | 80% | 69% | 59% | 78% | 64% | 56% | 73% |
| Orphan Designated | 13% | 9% | 16% | 6% | -1% | 12% | 9% | 3% | 14% | 18% | 12% | 25% |
| Proportion of Advice following Full submission | 78% | 73% | 83% | 81% | 71% | 92% | 80% | 73% | 88% | 75% | 68% | 82% |
| BNF1 cardiovascular | 10% | 7% | 14% | 17% | 6% | 27% | 9% | 3% | 14% | 8% | 4% | 13% |

| | **Total SMC sample (n=288)** | | | **SMC Recommended Technologies (n=54)** | | | **SMC Restricted Technologies (n=102** | | | **SMC Not Recommended Technologies (n=132)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | |
| system | | | | | | | | | | | | |
| BNF2 central nervous system | 21% | 16% | 26% | 15% | 5% | 25% | 19% | 11% | 26% | 26% | 18% | 33% |
| BNF3 ear, nose and oropharynx | 0% | 0% | 1% | 2% | -2% | 6% | 0% | 0% | 0% | 0% | 0% | 0% |
| BNF4 endocrine system | 8% | 5% | 12% | 9% | 1% | 17% | 10% | 4% | 16% | 7% | 2% | 11% |
| BNF5 eye | 1% | 0% | 2% | 4% | -1% | 9% | 1% | -1% | 3% | 0% | 0% | 0% |
| BNF6 gastro-intestinal system | 4% | 2% | 6% | 2% | -2% | 6% | 1% | -1% | 3% | 8% | 3% | 12% |
| BNF7 infections | 11% | 7% | 14% | 17% | 6% | 27% | 16% | 9% | 23% | 5% | 1% | 8% |
| BNF8 malignant disease and immunosuppression | 25% | 20% | 30% | 20% | 9% | 31% | 26% | 18% | 35% | 26% | 18% | 33% |
| BNF9 musculoskeletal and joint diseases | 3% | 1% | 6% | 0% | 0% | 0% | 5% | 1% | 9% | 4% | 0% | 7% |
| BNF10 nutrition and blood | 6% | 3% | 9% | 6% | -1% | 12% | 3% | 0% | 6% | 9% | 4% | 14% |
| BNF11 obstetrics, gynaecology, and urinary-tract disorders | 2% | 0% | 3% | 0% | 0% | 0% | 3% | 0% | 6% | 2% | -1% | 4% |
| BNF12 respiratory system | 4% | 2% | 6% | 2% | -2% | 6% | 3% | 0% | 6% | 5% | 1% | 9% |
| BNF13 skin | 4% | 2% | 6% | 7% | 0% | 15% | 5% | 1% | 9% | 2% | -1% | 4% |

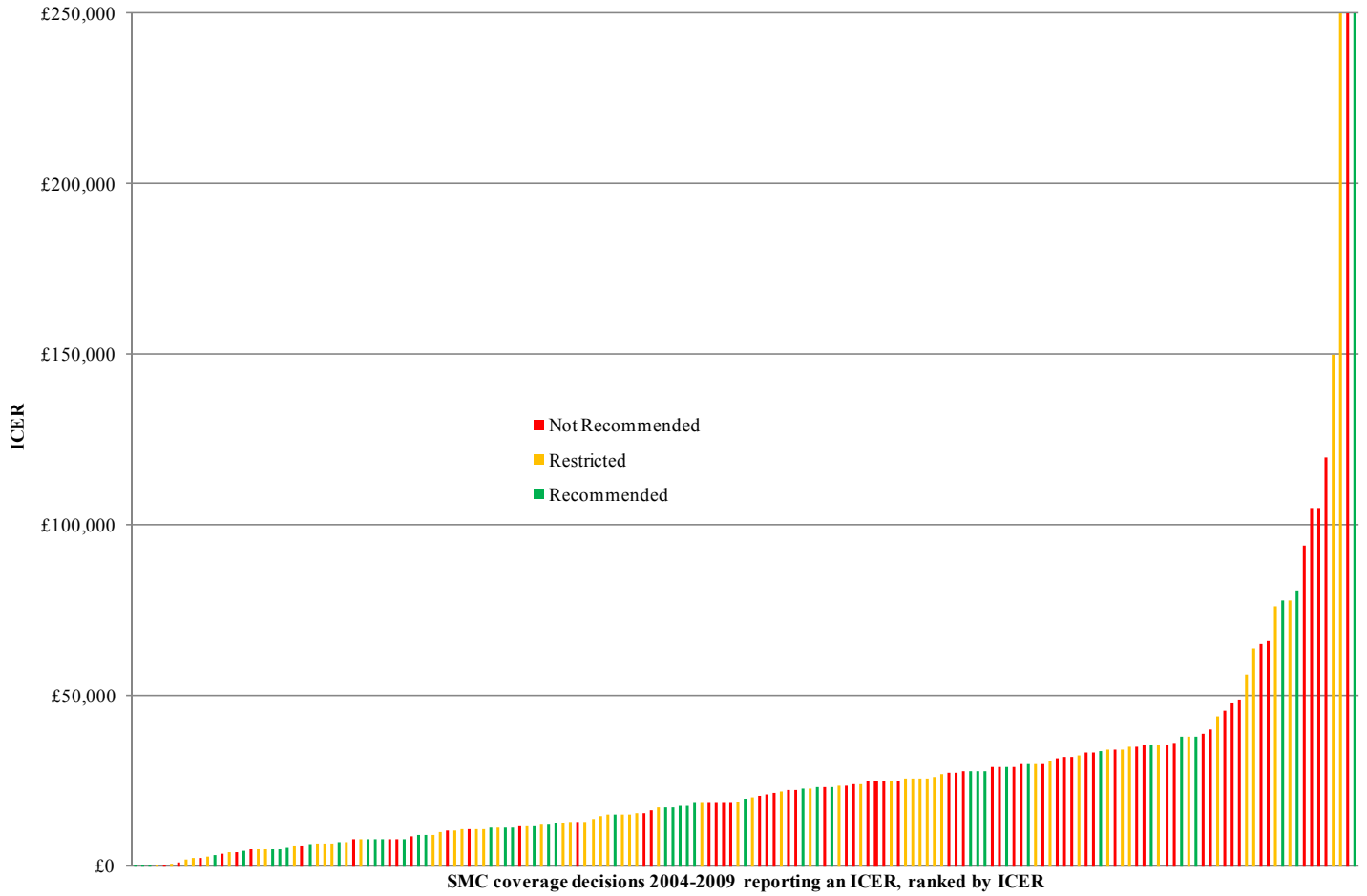
Economic Evidence

A range of economic related variables were included for analysis. 74% of SMC decisions were backed by use of CUA. The difference between recommended, restricted or not recommended interventions in the availability of non-CUA was not statistically significant (p=0.359). In almost all cases a single CUA, rather than multiple CUA, was reviewed.

For the interventions supported by a CUA, the ICER was significantly different between the recommended interventions (mean ICER of £11,893), compared to the restricted interventions (mean ICER of £26,316), and the interventions not recommended for use (mean ICER of £46,679) (p=0.0001). A descriptive analysis of the ICER within SMC advice is displayed in Figure 5.2. The appraisals are ranked by the size of the ICER, and the bars are coloured in accordance with the coverage decision made. In general, technologies that were recommended or restricted (green and orange bars) were found in the lower ICER ranges, while not recommended technologies (red bars) were located in the upper ICER range. However, the figure also shows that, while the ICER is an important variable in SMC decision-making, there are exceptions in its usefulness as a predictor of coverage decisions. There are recommended technologies with very high ICERs, as there are technologies with lower ICERs that are not recommended. This suggests the need to examine the role of a combination of factors to explain SMC decision-making.

Considerable effort was made to capture information on the uncertainty around base-case ICER estimates. This was done by recording results of probabilistic sensitivity analyses (the probability of ICER remaining below a set threshold), and by recording univariate sensitivity analyses (lowest ICER, highest ICER). The probability of the ICER remaining below £30,000 was 72% for recommended interventions, 71% for restricted interventions and 37% for interventions not recommended for use. The differences between restricted and not recommended interventions were statistically significant (p=0.018), and differences between all three coverage outcomes was statistically significant at the $p<0.10$ level.

**Figure 5.2 SMC: Base-case ICER reported in SMC appraisals, by coverage decision (n=169)**



With regard to univariate sensitivity analyses, the uncertainty around the base case ICER was measured by capturing information on the lowest ICER generated by the model and the highest ICER. The difference between these two provided useful information on the range of uncertainty around a base case ICER. The range of uncertainty was similar for those interventions that were recommended (£11,251-£31,647) or restricted for use (£8,963-£27,112), while the widest range of uncertainty was observed for interventions not recommended for use (£63,493-£147,174).

Aside from examining the results of CUA considered by the SMC review committee, information was also captured on whether alternative cost-effectiveness models (non-CUA) were considered in the decision-making process. 30% of interventions were supported by non-CUA models. Modest differences between decision outcome groups in the use of non-CUA models were not statistically significant (p=0.198).

The potential budget impact of a positive recommendation was analysed - the mean estimated maximum yearly budget impact across all decisions was in the order of £1.9 million. This ranged from £1.82 million for the recommended interventions to £1.21 and £0.90 million for the restricted and not recommended interventions, respectively. These differences observed between groups were not statistically significant.

SMC assessment process characteristics

A series of variables related to the decision-making process were recorded. Less than 1% of decisions formally considered a social perspective in the review process. In 41% of decisions, patient-group submissions were formally considered in the appraisal. On average the appraisal committee comprised of 25 members. In all of the decisions there was a cost-effectiveness evaluation component as well as a budget impact component. The Drug review process considers only one compound at a time. In terms of the submission process, 78% of submissions were “full submissions”, and 22% were “resubmissions”. None of the above variables differed significantly between outcome groups.

Socio-economic context of SMC decision-making

A series of variables were recorded to capture the socio-economic context in which SMC decisions were made. In particular, information was recorded on the year of the appraisal (mean 2007), size of population under SMC remit (approx. 5.14 million),

percentage of GDP spent on healthcare (8% on average across groups), whether appraisal coincided with an election year and whether the technology under appraisal was linked to a disease considered to be a 'priority' by Department of Health (66% of technologies appraised were directly linked to a priority disease area). There was strong correlation between the year of appraisal and the other socio-economic factors. The year of appraisal, population size, percentage of GDP spent on healthcare and mean healthcare expenditure per patient per year did not vary significantly between outcome groups.

Summary of descriptive analysis

Of the 36 explanatory variables explored within the SMC sample, descriptive analysis suggests that nine factors may play an important contributing factor in determining SMC decision-making (Table 5.4). For these variables, statistically significant differences were observed between interventions that were recommended, restricted and not recommended ($p \leq 0.05$). Highly statistically significant differences ($p \leq 0.01$) were observed for variables highlighted in bold in Table 5.4.

**Table 5.4 SMC descriptive statistics: statistically significant variables ($p \leq 0.05$)**

| **Evidence factors** | **Variables** |
|---|---|
| Clinical Package | **Size of RCT population**<br>Duration of RCT<br>**Active comparator used**<br>**Statistically significant superiority not demonstrated in RCT** |
| Economic Package | **ICER**<br>Uncertainty around ICER: probabilistic, **univariate** |
| Disease characteristics | **Prevalence of disease**<br>Alternative<br>**Orphan designated status**<br>BNF categories: eye disorders, gastro-intestinal system, infections |
| **Process factors** | - |
| **Socio-economic context factors** | - |

*Note: variables in bold text were statistically significant at the $p \leq 0.01$ level*

### *5.3.3 Multivariate Analysis*

Following the model specification process described in Chapter 3 which included the development of a preliminary model (Appendix B), the base case SMC regression model was developed which included seven variables (Table 5.5): the size and duration of the RCT, the ICER, if the technology was indicated for use in infections or skin diseases, number of patients eligible for treatment, and the availability of alternative therapies within NHS Scotland. When the coverage decisions were regressed with these

seven variables, the resulting pseudo R-squared was 0.10, suggesting that these seven variables explained approximately 10% of the variability in SMC coverage decisions. The ICER and number of patients eligible for treatment were the only variables that had statistically significant impact on coverage decisions whether between a recommendation and a restriction or recommendation and a non-recommendation. A unit increase in the ICER increased the probability of a decision to restrict or not recommend – which was statistically significant in both arms of the model (p=0.33 and p=0.018, respectively). A unit increase in the number of patients eligible for treatment statistically significantly decreased the odds of both a restriction and recommendation.

**Table 5.5 Multivariate analysis of SMC coverage decisions 2005-2009: base case model results**

| Restricted Technologies | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| RCT size | 0.000056 | 0.42 | -0.00008 | 0.00019 |
| RCT duration of follow-up | -0.0039 | 0.19 | -0.0098 | 0.0019 |
| ICER | 0.000025 | 0.033 | 0.000002 | 0.0000472 |
| Infectious Diseases | -0.16 | 0.74 | -1.10 | 0.78 |
| Skin Diseases | -0.32 | 0.66 | -1.74 | 1.10 |
| No. of patients eligible for treatment | -0.000016 | 0.083 | -0.0000337 | 2.08E-06 |
| Presence of alternative therapy | -0.53 | 0.39 | -1.73 | 0.67 |
| Constant | 0.97 | 0.15 | -0.35 | 2.29 |
| Not Recommended Technologies | Log Odds | P value | 95% Conf. Interval | |
| RCT size | -0.00029 | 0.067 | -0.00060 | 0.00002 |
| RCT duration of follow-up | -0.0096 | 0.008 | -0.017 | -0.0025 |
| ICER | 0.000027 | 0.018 | 0.0000048 | 0.000050 |
| Infectious Diseases | -1.67 | 0.004 | -2.81 | -0.53 |
| Skin Diseases | -1.72 | 0.061 | -3.52 | 0.08 |
| No. of patients eligible for treatment | -0.00002 | 0.087 | -0.00003 | 0.00000 |
| Presence of alternative therapy | -0.98 | 0.106 | -2.16 | 0.21 |
| Constant | 2.18 | 0.001 | 0.87 | 3.49 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.10.

The impact of the remaining variables varied according to the nature of the decision. When the decision was between a recommendation and a non-recommendation, several clinical and disease variables were found to be significant. The size and duration of the RCT, if the technology was indicated for the treatment of infections or skin diseases, and the presence of an alternative therapeutic option, all had a statistically significant impact on the probability of a recommendation versus a non-recommendation. A unit increase in the number of patients enrolled in the RCTs included for review within the appraisal process decreased the log odds of a non-recommendation relative to a recommendation (p=0.067). Similarly, the duration of follow-up within RCTs reviewed by the SMC appeared to have an impact on coverage decisions – a unit increase in duration increased the probability of a decision recommendation (p=0.008). The nature

of the disease for which the technology was indicated, particularly if it was indicated for treatment of infectious or skin diseases, appeared to decrease the odds of a non-recommendation (p=0.004, p=0.061 respectively). Perhaps counter-intuitively, the presence of a pre-existing alternative therapy available within the Scottish NHS increased the odds of a recommendation.

Impact of alternative model specifications- sensitivity analyses

Sensitivity analyses were conducted on the SMC regression model, as described in Chapter 3. This included i) examining the impact of a binary rather than three-category outcome variable; ii) restricting the base case analysis to complete observations, thus excluding observations with imputed values, and (iii) examining the impact of assuming ordinality of the outcome variable.

A first sensitivity analysis was conducted on the SMC regression model by collapsing the three category outcome variable into a binary outcome variable. A logistic regression was performed examining the log likelihood and the odds of coverage versus non coverage. In this binary model, strong interaction was observed between the ICER and the variable signalling the presence of an alternative therapy. When the latter variable was removed from this model the remaining variables, as shown in Table 5.6, maintained a similar direction of effect and statistical significance of these effects. An increase in the sample size of the RCT was associated with an increase in the odds of a recommendation, as was a unit increase in the duration of the clinical trial. The ICER continued to have a statistically significant effect, as observed in the base case analysis. Technologies indicated for the management of infectious or skin diseases appeared to have increased odds of being covered. Finally, a unit increase in the prevalence of the disease also increased the odds of coverage. Overall, this sensitivity analysis demonstrated that, through the use of a binary outcome variable as opposed to a three-category outcome variable, the variables that were associated with a statistically significant impact on SMC coverage decisions in the base case analysis had a similar effect in this alternative binary outcome model. However, it should be noted that one variable had to be dropped due to the interaction observed which had not been found in the base case analysis. In addition, the use of a binary modelling approach was not able to reflect how the role of each factor varies according to the decision being made: i.e. the impact of a single variable on the decision to recommend or not recommend may be different than its impact on the decision to recommend or to restrict. This was clearly

shown to be the case in the base case model where many of the factors had a statistically significant impact when the decision was between recommendation and non-recommendation, but not between recommendation and restriction.

**Table 5.6 Sensitivity Analysis 1. Multivariate analysis of SMC coverage decisions 2005-2009: sensivitiy analysis using binary outcome variable (covered vs. not covered)**

| | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| RCT size | 0.000355 | 0.049 | 1.94E-06 | 0.000708 |
| RCT duration of follow-up | 8.06E-03 | 0.013 | 0.001688 | 1.44E-02 |
| ICER | -7.84E-06 | 0.076 | -1.7E-05 | 8.35E-07 |
| Infectious Diseases | 1.608167 | 0.005 | 0.475342 | 2.740993 |
| Skin Diseases | 1.587428 | 0.049 | 0.00466 | 3.170196 |
| Disease prevalence | 7.38E-06 | 0.104 | -1.51E-06 | 1.63E-05 |
| Constant | -0.49728 | 0.02 | -0.91499 | -0.07957 |

A second sensitivity analysis was performed in which the sample of analysis excluded incomplete observations. This led to a reduction in sample size from 287 to 143 observations. The results of the analysis are provided below, and show that the number of variables that have a statistically significant impact on SMC coverage decisions is reduced to five variables: duration of RCT follow-up, ICER, if the technology is indicated for the treatment of infectious diseases, and disease prevalence (Table 5.7). The pseudo R-squared value was 0.14, thus this model explained a higher proportion of the variability observed in SMC coverage decisions compared to the base case model (pseudo R-squared of 0.10). However, the results of the model show that the exclusion of appraisals with missing data alters three of the seven variables that demonstrated significance in the base-case model: the impact of sample size of the RCT and of technologies treating skin diseases was no longer observed, nor was the impact of the presence of alternative therapies observed in this sensitivity analysis. The remaining variables in the model appear to maintain a similar direction of effect as that observed in the base case analysis. However, of note is the fact that infectious diseases technologies and disease prevalence appear to have a statistically significant impact on both the log odds of restriction and non-recommendation, while in the base case analysis, a statistically significant impact was only observed in the latter instance. Thus, the exclusion of incomplete observations in this sub-analysis does not change the importance of the ICER, the duration of the RCT, the prevalence of the disease and the nature of the disease (i.e. if indicated for infectious disease) on SMC coverage decisions. These four variables that were found to be important in the base case analysis remain as such. However, three variables which had significant impact in the

base-case analysis no longer show such an effect in this sub analysis. This may reflect the bias created in the sample by the removal of incomplete observations.

**Table 5.7 Sensitivity Analysis 3. Multivariate analysis of SMC coverage decisions 2005-2009: sensivitiy analysis excluding incomplete observations from the sample**

| | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Restricted technologies | | | | |
| RCT duration of follow-up | -0.0011 | 0.0042 | -0.27 | 0.79 |
| ICER | 0.000052 | 0.000025 | 2.10 | 0.036 |
| Infectious Diseases | -1.97 | 0.83 | -2.39 | 0.017 |
| Disease prevalence | -0.000031 | 0.000017 | -1.82 | 0.069 |
| Constant | 0.60 | 0.51 | 1.17 | 0.243 |
| Not Recommended Technologies | | | | |
| RCT duration of follow-up | -0.01 | 0.01 | -2.66 | 0.008 |
| ICER | 0.000056 | 0.000 | 2.30 | 0.021 |
| Infectious Diseases | -1.69 | 0.76 | -2.21 | 0.027 |
| Disease prevalence | -0.000073 | 0.000039 | -1.87 | 0.062 |
| Constant | 1.45 | 0.51 | 2.82 | 0.005 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared = 0.14

In the third sensitivity analysis, ordinality of the outcome variable was assumed. This is in contrast with the base case analysis, where ordinality was not assumed and multinomial logistic regression was used. In this sensitivity analysis ordinal logistic regression was used. The detailed results of this analysis are provided in Appendix B, provide generally very similar results to the base case analyses run using a multinomial logistic regression model.

## 5.4 Discussion

The overall objective of this chapter was to examine the evidence, process and context factors that influence decisions made by SMC to accept for routine use, accept for restricted use or not recommend new technologies for use in NHS Scotland. In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, and in light of evidence review presented in Chapter 2, A wide range of explanatory variables were included in the analysis, reflecting clinical and economic characteristics of the technology under appraisal, the appraisal process itself and the socio-economic context in which the SMC operates. In addition to the general aims of the research, specific hypotheses relevant for SMC decision-making were explored and are discussed below.

To help evaluate the internal validity of the multivariate analyses performed, the base case model results were shared for review with a member of the SMC (Interview with

Dr. A. Walker[1]). The aim of this interaction was to ascertain if the SMC characteristics were accurately captured in the sample used for the analysis, if the approach to the analysis was clear and, in particular, the reaction to the model results and potential for suggestions or additional analyses.

### *5.4.1 Pattern of SMC coverage decisions*

The analysis of SMC decisions involved the review of 288 technology appraisals. The most common coverage decision by the SMC was to not recommend funding for the new technology (46%), followed by restriction of funding (35% of technologies), and technologies recommended for funding (19%). Multivariate analysis of SMC coverage decisions suggest that seven variables appear to have a significant effect on coverage decisions: the sample size and duration of the RCT(s); the incremental cost-effectiveness ratio (ICER); if the technology was indicated for use in infections or skin diseases, the prevalence of the disease in question, as well as whether there was an alternative therapy available within NHS Scotland. Other analyses of coverage decisions made by the SMC provide different proportions of coverage types, although this appears to be due to differences in time horizon. For example, the SMC Annual Report (SMC 2008) summarises coverage decisions for 2008: 31% accepted for use, 36% of technologies accepted for restricted use, and 33% technologies not recommended for use. In a similar exercise but looking at decisions in the period 2007-2009, Kanavos et al. (2010) find that the SMC recommended 28% of technologies, restricted 40% of them and did not recommend 32% of technologies. Thus, the period used within the analyses produced in this thesis, spanning 2005-2009, has a higher proportion of non-recommendations in its sample than observed in other publications reporting SMC decision-making. Dr. A. Walker[1] suggested that such differences can be attributed to the use of different time horizons (2005-2009 vs. 2007-2009 or single years) and also to the fact that publications report SMC coverage decisions across all types of submissions, whether full submissions, resubmissions, abbreviated submissions or IRPs. Within the SMC sample used for this thesis, only full or re-submissions were included.

---

[1] Walker, Andrew. Senior Lecturer in Health Economics, University of Glasgow and member of the New Drugs Committee at the Scottish Medicines Consortium. Interviewee: Karin Cerri. Interviewed by telephone, on December 20th 2010. Meeting minutes are provided in Appendix E.

### *5.4.2 The role of clinical evidence factors in SMC decision-making*

The duration and size of the RCTs supporting the technology had a significant impact on coverage decision. Specifically, increasing trial duration and trial sample size was associated with decreased log odds of a non-recommendation versus recommendation, and a decreased log odds of restriction relative to recommendation. Technologies recommended by the SMC tended to be supported by on average larger trials of longer duration (sample of 1,532 patients, with trial duration of 60 weeks), compared with technologies that were restricted (mean sample of 1360 patients, with trial duration of 50 weeks) or not recommended (mean sample of 484 patients, with mean trial duration of 35 weeks) by SMC.

Disease characteristics were also found to play an important role in SMC coverage decisions. The nature of the disease and the size of eligible patient population were found to have a statistically significant effect. Perhaps unexpectedly, an increasing size of eligible population of the disease targeted by the technology was found to decrease the odds of non recommendation relative to recommendation. This result is contrary to the hypothesis that there would be a relationship between budgetary impact of new technologies, which are in part driven by the anticipated volume of up-take of a new technology, and SMC decision-making, specifically that an increase in budgetary impact would be associated with a decrease in the log odds of recommendation relative to restriction and non-recommendation. This data however suggests that an increase in patient volume, which would tend to increase budgetary impact, is not associated with decreased log odds of recommendation – but rather the contrary. The mechanism for this was suggested to be related to the characteristics of those technologies that have low number of patients eligible for treatment (Dr. A. Walker). When appraisals were grouped by the size of the eligible population for which the technology was indicated (500 patients or less per year "low eligible population", vs. >500 patients per year), the mean ICER in the low eligible population group was more than twice that of the technologies with larger eligible population (£48,480 vs. £19,945). The technologies for which there are an estimated <=500 eligible population appeared to be supported by lower quality set of supportive clinical evidence relative to the >500 eligible population group (Table 5.8)

**Table 5.8 SMC coverage decisions – comparison of characteristics of evidence supporting technologies, by size of eligible population**

| Size of eligible population | ICER | No. Of RCTs | RCT sample size | RCT duration (Weeks) |
|---|---|---|---|---|
| <= 500 eligible patients per year) | £48,480 | 1.5 | 423 | 46 |
| >500 eligible patients per year | £19,945 | 2.7 | 1,753 | 42 |

Technologies indicated for the management of infectious diseases increased the odds of a recommendation relative to a non-recommendation. Among the set of technologies for infectious diseases (n=28), 35% were for the treatment of HIV infection, while the majority (54%) of these technologies were antibiotic or anti-fungal agents (n=15). Among this set of technologies (n=15), none of the appraisals reported ICERs, and the clinical data supporting these studies was lower than the average for the entire SMC sample. Specifically, antibiotic/anti-fungal agents were supported by 1.4 studies (vs. mean of 2.2) and approximately half of the mean sample size (999 patients in total SMC sample vs. mean 513 patients in this subsample). Despite the relatively low quality supporting evidence, 80% of the antibiotic/anti-fungal agents were restricted (12 of 15) and only 2 were not recommended. A plausible explanation for the impact of infectious diseases on coverage decisions could be the fact that antibiotics are included within this BNF category, and that on the one hand, the quality of data for antibiotic therapy is not very high, but on the other hand the concern over developing resistance to antibiotics suggests that there is a need to allow for new antibiotics to be available for use (Interview with Andrew Walker, 2010). This may explain the result obtained in the model with effects which are significant when there is a trade-off between recommendation and non-recommendation, but not statistically significant when the trade off between restriction and non-recommendation.

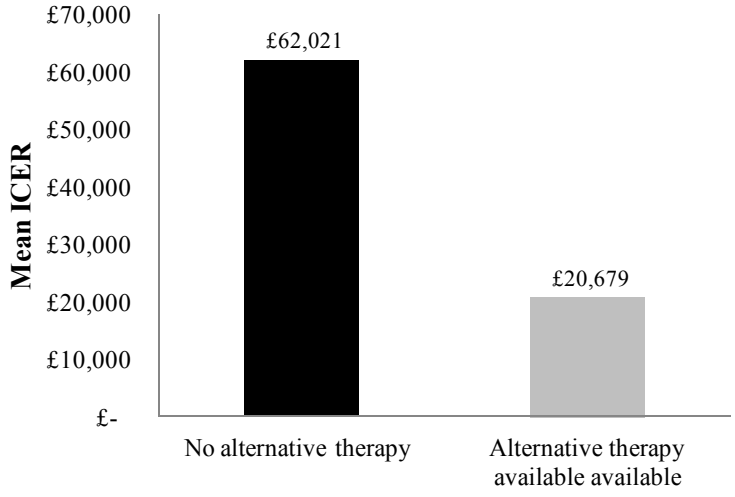
### *5.4.3 The role of economic evidence in SMC decision-making*

The ICER was found to have a significant positive effect on the odds of a restriction and non-recommendation relative to a recommendation. This can be seen to reflect the SMC's objective to identify those technologies that represent value for money. As shown in Figure 5.2, technologies that were recommended or restricted were found in the lower ICER ranges, while not recommended technologies were located in the upper ICER range. However, the figure also shows that, while the ICER is an important variable in SMC decision-making, there are exceptions in its usefulness as a predictor of coverage decisions. There are recommended technologies with very high ICERs, as

there are technologies with lower ICERs that are not recommended. The role of uncertainty around the ICER was challenging to assess due to poor reporting of information. Of the 288 observations, uncertainty information (whether univariate or probabilistic) was missing in more than 200 observations.

The results also suggest that in its decision-making the role of the ICER has a prominent impact, relative to other factors such as degree of unmet medical need. The presence of an alternative therapy indicated in the same population to the technology under evaluation appears to impact significantly on SMC coverage decisions, although perhaps not in the direction that had been hypothesised based on the literature. The analysis suggests that the presence of an alternative therapy decreases the odds of a non-recommendation (p=0.106). In other words, those technologies where no alternative was present had increased odds of non-recommendation, relative to those technologies where an alternative was already available in NHS Scotland, all other things being equal. It is noteworthy that about 40% of those technologies with no alternatives were orphan-designated technologies. Moreover, those technologies where no alternative was available also had higher mean ICERs compared to those technologies where alternatives were available – almost three times as high (£62,021 vs. £20,679 respectively, Fig. 5.3). This provides an example of the various clinical, disease and economic factors intertwined within an HTA process. In addition, these results raise questions about the extent to which the SMC encourages innovation in the pharmaceutical industry.

**Figure 5.3 SMC: mean ICER for technologies where alternative therapy is present or absent**



### *5.4.4 No clear impact of process or socio-economic context variables on SMC decision-making*

It was hypothesised that SMC decision-making would be impacted by the process through which the technology was appraised. In particular it was hypothesised that the type of submission (whether first-time submission or re-submission) would impact on decision outcomes. However, this was not found to be the case – across the outcome categories, there were no statistically significant differences in the proportion of technologies appraised under first or re-submissions. Indeed, none of the variables collected to assess impact of process on decision-making (including use of economic evidence as part of process, size of committee) were associated with significant impact on the outcome variable. With regards to the socio-economic context variables – none were found to have a statistically significant and stable effect in the multivariate analyses nor in the descriptive analyses. This evidence therefore is not in support of the hypothesised change in decision-making outcomes over time.

### *5.4.5 Limitations*

When examining the results of the multivariate analyses, there are several limitations that need to be taken into account. The first is that the SMC, compared with other agencies like NICE, provides relatively limited information in the public domain on the evidence reviewed and considered in their decision-making process. In general, the SMC Advice reports that are publicly available provide concise summaries of key issues but do not document details on the various clinical considerations or economic arguments to which they were exposed; primarily they highlight those considerations that were found to be drivers of their decision. On the one hand, this helps the data extraction process by providing the key data that was felt, by the agency, to drive their decision-making. On the other hand, the aim of this thesis and analysis was to collect as much objective evidence as possible on factors driving decision-making. The lack of detail in reporting led to higher rates of incomplete observations, particularly with regard to information on the uncertainty around incremental cost-utility/effectiveness ratios, compared to agencies like NICE where a larger quantity of information is publicly available, including manufacturer submissions (depending on the appraisal process used). The lack of data linked to this reporting style was managed by using imputation techniques in the multivariate analysis. The implications of using such techniques, versus restricting the analysis to complete observations were assessed in a sensitivity analysis conducted on the sample of coverage decisions for which the data

was complete. The results of these sensitivity analysis showed that limiting the analysis to complete observations, and thereby removing 130 decisions from the 288 SMC advice reports, may lead to bias in the sample for analysis. Despite this bias, four of the seven variables significant in the base-case analysis remain significant in this sub-analysis, thus confirming their important role in SMC coverage decisions.

An important factor to take into account when examining SMC coverage decisions is its reliance on manufacturers' submissions to formulate its advice. There is no third party or significant additional new analysis performed on the evidence submitted by the manufacturer. The focus of the SMC is to critically review the submitted evidence in order to reach a conclusion about the degree of certainty or uncertainty around the effects and value for money of the technology under appraisal. Given the lack of accessibility to manufacturer submissions in the public domain, it was not possible to take into account in the analyses to what degree the SMC advice was driven by the manufacturer submission strategy relative to SMC decision-making criteria (Conversation with Andrew Walker, 2010). For example, for a technology that was accepted for restricted use, it was not possible to ascertain if this restriction was proposed and implemented by the SMC, or whether the restriction was proposed by the manufacturer in their submission. Nevertheless, while there was lack of complete information on the evidence and manufacturer strategy used within SMC assessments, it can be argued that whether a particular coverage decision is proposed by the manufacturer or by the SMC does not hamper the ability to identify the effect and significance of the clinical, economic and disease characteristics of the technologies assessed by the SMC on its coverage decisions.

It is noteworthy that none of the appraisal process characteristics or socio-economic factors appeared to have a significant effect on coverage decisions. This suggests the general stability of the appraisal process and socio-economic context of decision-making, and therefore the lack of significant variation within the period for which coverage decisions were extracted. In addition, it should be noted that socio-economic factors varied at the agency level, rather than at the decision-level: thus the degree of variability was substantially reduced to annual changes in socio-economic factors. The relative importance of socio-economic and process factors will be examined further in the pooled analysis of coverage decisions from all four HTA bodies, presented in Chapter 8.

In summary, the overall objective of this chapter was to examine the factors that influence decisions made by SMC to recommend, restrict or not recommend pharmaceutical technologies for use in its respective healthcare systems, with a focus on research hypotheses specific to SMC decision-making. The results suggest that the variability in coverage decisions observed can be explained by a combination of clinical, disease and economic factors. It is noteworthy that the role and effect of the majority of variables differed according to the nature of the decision. This suggests that the use of a multinomial outcome variable facilitates the ability to identify potential differences in the impact of the same factor on different coverage decisions, which would not have been possible with a binary outcome category. The analyses did not support the hypothesis that process factors impact on SMC decision-making, nor that socio-economic factors, measured by time, significantly impact on SMC decision-making. The analysis confirmed the important role of the ICER in SMC decision-making in which increases in the ICER are associated with decreasing log odds of recommendation relative to restriction or non-recommendation. The results also suggest that the ICER may take a central role in decision-making over and above other considerations such as unmet medical need. This was observed in a closer examination of technologies for which there were no alternative regimens available – lack of alternative tended to increase rather than decrease the odds of non-recommendation contrary to the hypothesised effect. Contrary to expectations, budgetary impact was not found to play a significant role in SMC decision-making, based on the multivariate model generated.

## 5.5 References

Association of the British Pharmaceutical Industry (ABPI). 2010. "Facts & Statistics from the pharmaceutical industry - Medicines and the NHS" http://www.abpi.org.uk/statistics/section.asp?sect=4#15. Viewed 28 January 2010.

BBC News.2005. "2005: Historic third term for Labour" http://news.bbc.co.uk/2/hi/uk_news/politics/vote_2005/6994476.stm. Viewed 22 December 2010.

———.2007. "Another chapter ends at Holyrood" http://news.bbc.co.uk/2/hi/uk_news/scotland/6512689.stm. Viewed 28 June 2009.

Cairns, J. 2006. Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy* 76: 134–143.

Drummond, M., and A. Mason. 2009. Rationing new medicines in the UK. *BMJ* 338:a3182.

European Medicines Agency. 2009. "Adalimumab: European Public Assessment Report". http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000481/WC500050870.pdf. Accessed 5 January 2011.

———. 2011. European public assessment reports. http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&murl=menus/medicines/medicines.jsp&mid=WC0b01ac058001d125. Viewed between 1st of June 2009-30th March 2010.

Information Services Division (ISD). 2009. "Summary cost statistics 2009". http://www.isdscotland.org/isd/6480.html. Viewed 28 June 2009.

Joint Formulary Committee. 2010. British National Formulary. 60 ed. London: British Medical Association and Royal Pharmaceutical Society. http://bnf.org/bnf/index.htm

National Institute for Health and Clinical Excellence. 2008. NICE technology appraisal guidance 143 - Adalimumab, etanercept and infliximab for ankylosing spondylitis. London: National Institute for Health and Clinical Excellence, 2008.

National Office for Statistics. 2009. "Key demographic and health indicators, 1976 onwards: Population Trends. ONS 1976-2008 data". http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=9537 . Viewed 23 June 2009.

NHS Health Scotland. 2004. Working in Partnership to improve health and reduce inequalities; Business Plan 2004-2005. Glasgow: NHS Health Scotland. http://www.healthscotland.com/uploads/documents/BDP3604-full.pdf

———. 2005. Health Scotland Annual report 2005/2006. Glasgow: NHS Health Scotland. http://www.healthscotland.com/uploads/documents/2682-Annual%20Report_2362_11_2006.pdf.pdf

———. 2006. Health Scotland Delivery Plan 2006 – 2007. Glasgow: NHS Health Scotland.http://www.healthscotland.com/uploads/documents/13792_report_combined.pdf

———. 2007. Better Health, Better Care: Action Plan: What It Means For You. Glasgow: NHS Health Scotland. http://www.scotland.gov.uk/Resource/Doc/210642/0055693.pdf

———. 2008. NHSScotland Chief Executive's Annual Report 2007/08, Appendix B HEAT Targets 2008/09. Glasgow: NHS Health Scotland. http://www.scotland.gov.uk/Publications/2008/11/28081831/9

OECD. 2009. "OECD Health Data - Version: November 09". http://www.ecosante.org/index2.php?base=OCDE&langs=ENG&langh=ENG. Viewed 22 June 2010.

Scottish Medicines Consortium. 2006. Adalimumab 40mg pre-filled syringe (Humira®) (No. 300/06). http://www.scottishmedicines.org.uk/files/adalimumab_40mg_Humira_300_06.pdf

———. 2010a. Introduction and availability of newly licensed medicines in the NHS in Scotland. http://www.scottishmedicines.org.uk/files/CEL2010_17.pdf

———.2010b. Guidance to Manufacturers for Completion of New Product Assessment Form (NPAF) (Revised June 2010). http://www.scottishmedicines.org.uk/Submission_Process/Submission_Guidance_and_Templates_for_Industry/Templates-Guidance-for-Submission/Templates-Guidance-for-Submission

———. 2011. "SMC Meeting Minutes" http://www.scottishmedicines.org.uk/About_SMC/Minutes/Minutes. Viewed 28 June 2009.

# 6 Empirical analysis of CVZ coverage decisions

Adalimumab was reviewed by NICE, SMC, CVZ and HAS for the treatment of ankylosing spondilitis. The CVZ provided the following guidance:

> "...[adalimumab is] indicated for an insured member of eighteen years of age or older (...) with severe active ankylosing spondylitis where there is evidence of insufficient response to at least two non-steroidal anti-inflammatory drugs at maximum dose and other conventional therapy". (CVZ 2006 section 7a)[1].

Based on its assessment of adalimumab, the CVZ restricted its use to those patients that had experienced 'insufficient' response to at least two previous therapies. This decision is similar to that made by both NICE and the SMC, despite the fact that the Dutch reimbursement system and CVZ assessment process differ from the pharmaceutical funding mechanisms within the NHS structures in Scotland and England and the methods by which NICE and SMC appraise technologies. To what extent do evidence, process and context factors play a role in CVZ decision-making? And are the factors that were found to impact significantly on NICE and SMC decision-making also reflected within CVZ coverage decisions?

This chapter provides an empirical analysis of coverage decisions made by CVZ in 2004-2009. First, an overview of the CVZ, its objectives and appraisal process is provided. The methods for the analysis are then outlined, building upon the methods discussed in Chapter 3. Following this, the results of the descriptive and multivariate analyses of the CVZ coverage decisions are reported and explored, and limitations considered. The chapter concludes with a brief discussion on the empirical analyses performed for the CVZ.

## 6.1 CVZ Appraisal Process

In the Netherlands, the College Voor Zorgverzekeringen (CVZ) has an important role in supporting and maintaining the quality, accessibility and affordability of health care in the Netherlands. It has both an advisory and implementation role for two social health insurance funds: the Zorgverzekeringswet (Zvw) and the Algemene Wet Bijzondere

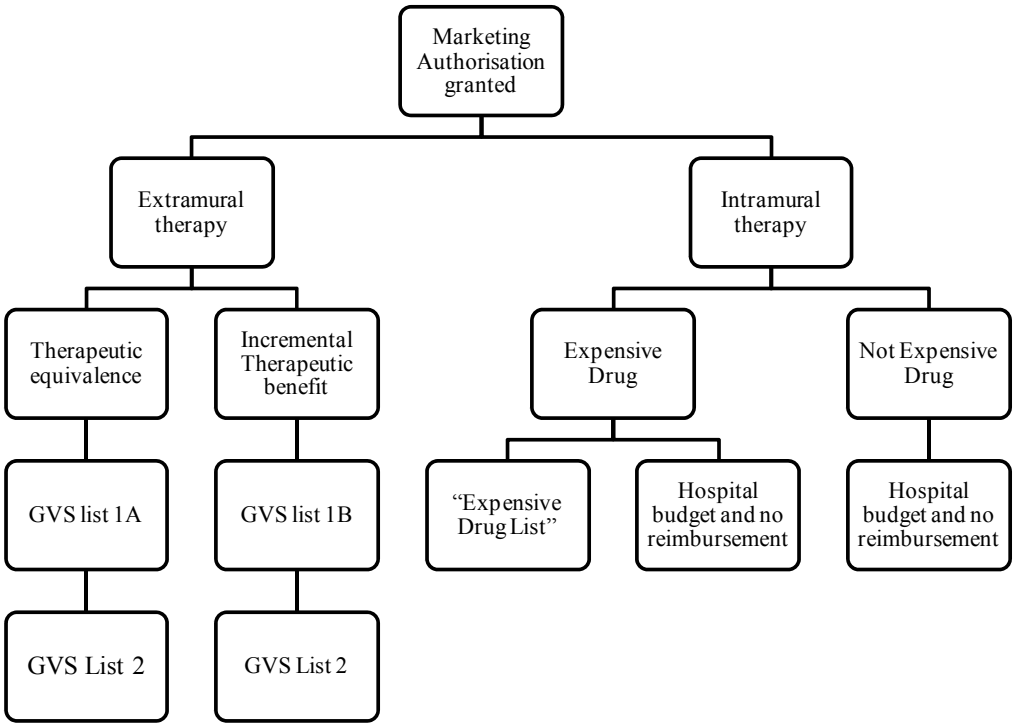
[1] Original CVZ statement: "uitsluitend voor een verzekerde van achttien jaar of ouder (...) met ernstige actieve spondylitis ankylopoëtica waarbij er sprake is van onvoldoende respons op ten minste twee prostaglandinesynthetaseremmers in maximale doseringen en andere conventionele behandeling." CVZ 2006 section 7a.

Ziektekosten (AWBZ). The Commissie Farmaceutische Hulp (CFH) is a part of the CVZ. It is tasked with the assessment of the therapeutic value of new technologies for inclusion in the Medicine's Reimbursement System (Geneesmiddelen Vergoedings Systeem – GVS), or inclusion in the various reimbursement policies of the Dutch Health Authority (Nederlandse Zorgautoriteit - NZa).

The CFH is composed of 20 external members, with a maximum eight-year length of service that meets on a monthly basis to review applications for reimbursement (CFH 2011). Unlike NICE and its MTA appraisal process, there is no third party evaluation system utilised by the CFH. Collectively, the role of the CFH and CVZ is to advise the Minister of Health on whether to include the technology for reimbursement and if so, on what basis. Thus, the Ministry of Health has the final decision-making authority with regard to the reimbursement of new technologies.

The GVS includes the list of medicines that are reimbursed under the conditions of the Zorgverzekeringswet (Zvw). The technologies included in the GVS list are reimbursed by all the sickness funds. Those that are included on the special policy lists (e.g. expensive drug list) are reimbursed by the Ministry of Health, Welfare and Sports (CVZ 2010a). There are two lists within the GVS: the 1A and 1B list (CFH 2010a; Sorenson et al. 2008; Stolk and Rutten 2005). Figure 6.1 summarises the coverage options available within the Dutch health care system. List 1A is for technologies used to treat a similar patient population with comparable therapeutic value and mode of administration. Technologies in List 1A are 'clustered' around one reference price for each 'basket' of technologies. List 1B includes medicines that are not clustered and thus do not fall within the reference pricing system. In addition, there is a List 2 which includes medicines for which there are restrictions on use (CVZ 2010a). Unlike technologies on List 1A which have a reimbursement limit calculated based on the prices of other technologies within the cluster, technologies on 1B and 2 do not have a pre-specified limit. For those technologies in List 1A where the acquisition costs are higher than the limit, the patient is asked for a co-payment by the health insurers to be able to access the medications.

**Figure 6.1 Pharmaceutical Reimbursement System in the Netherlands**




In its review of the reimbursement request by the manufacturer, the CVZ aims to provide advice to the Ministry of Health on two aspects: firstly, the positioning of the technology under appraisal vis a vis existing technologies; and secondly, under which reimbursement list should the technology be funded – the GVS 1A or 1B list, or specific reimbursement policies (e.g. the expensive drug list, List '2'). The lists differ in the appraisal process and criteria used. For inclusion on the GVS 1B list, since 2005 pharmacoeconomic evaluation is a mandatory requirement (Stolk and Rutten 2005). For inclusion in the expensive drug list, pharmacoeconomic evaluation and a protocol for future pharmacoeconomic evaluation based on real-life effectiveness evidence is needed. Specific types of technologies are exempt from pharmacoeconomic evaluation, including orphan designated technologies, technologies for which the total budgetary impact of the introduction of the technology after a five year period is lower than EUR €500,000, or technologies for which there is equivalence in both effectiveness profile and price.

In assessing the positioning of the technology, the therapeutic value of the technology plays an important role in the CVZ recommendation. The CVZ considers three levels of therapeutic value in comparison to a standard of care already reimbursed: i) lower therapeutic value (e.g. worse safety profile, lower efficacy), ii) comparable therapeutic

benefit (e.g. no relevant differences in favour or against the technology have been identified); or iii) added therapeutic benefit (where there are relevant advantages in either efficacy, safety or both) (CVZ 2010b).

In order to assess into which therapeutic benefit category the technology should be placed, the CVZ takes into account several factors, including the disease for which the technology is indicated and whether or not there is a standard of care already reimbursed for that system with the same indication, based on clinical guidelines and clinical criteria. The review then considers various aspects of the technology to determine whether an incremental therapeutic benefit exists. This includes the review of efficacy and safety, based on randomised clinical control trials (RCTs) and safety registries, as well as information about the experience with the technology - such as time on the market - to assess the risk of unknown side effects occurring over time and to increase the certainty around the therapeutic benefit of the technology. The CVZ evaluates the suitability of the technology – that is, if there are contra-indications for specific subgroups of the population e.g. the elderly or for paediatric use. The CVZ also reviews the 'ease of use' of the technology – including aspects such as dose frequency, whether it is intravenous or oral, how it is packaged and any other aspect that could influence the ease of use for patients or carers administering the technology. The financial implications or cost-consequences of adoption of the technology are also assessed, similar to budget impact estimation.

Once all of the above aspects have been considered for each individual technology, the importance of each factor is weighted and technologies are compared, to come to a decision on the category of therapeutic benefit that should be applied (lower benefit, comparable benefit or higher benefit). The CVZ states that in this weighting process, several factors are taken into account, including the severity of the disease, whether the disease is chronic, and whether alternative therapies are available (CFH 2010b). Once the CVZ has provided its advice to the Ministry of Health, and the Ministry of Health gives its formal approval, the reimbursement decision for extramural[2] medicines is published on the Pharmaco-therapeutic Compass (*Farmaco-therapeutisch Kompas)* (CFH 2010a).

[2] Extramural medicines are those medicines prescribed outside of a hospital inpatient setting.

A special process exists for so called 'expensive intramural medicines' (*Beleidsregel dure geneesmiddelen*) which are restricted for use in inpatient hospital conditions (College Tarieven Gezondsheidszorg ZorgAutoriteit 2006). A technology can be considered for inclusion in this list if all of the following conditions are met: i) the total costs for the technology are equal to or higher than 0.5% of the total hospital pharmaceutical budget at national level; ii) if the technology appears to have a therapeutic benefit that is of value to the healthcare system; and iii) if there is an agreement in place on the data and research that will be carried out after reimbursement to establish the true effectiveness of the technology in real-life.
Three years after inclusion in the expensive drug list, the CVZ re-appraises the technology and evaluates its true demonstrated effectiveness, providing advice to the NZa on whether to maintain or remove the technology from the list. This process, in fact, represents a type of conditional coverage mechanism, where the technology is introduced, observed and re-evaluated in a defined period of time according to a defined process. This system was put in place in 2006.

The CVZ also has a particular role, perhaps less common in other HTA bodies, of reviewing technologies for unlicensed indications, upon the request of the health insurance bodies. In this situation, the CVZ is asked to establish if the unlicensed indication is rare (less than 1:150,000 population), if there is a scientific basis for the efficacy of the technology in this unlicensed indication and if there is no other alternative therapy available in the Netherlands for the condition under review.

## 6.2 Methods

The overall objective was to examine the factors that drive decisions made by CVZ to recommend, restrict or not recommend new technologies for use in the Dutch healthcare system. In addition to the general analysis aims described in Chapter 1, the particular hypotheses relevant for the modeling of coverage decisions

Box. 6.1 In light of the discussions presented in Chapter 2, CVZ-specific research objectives were to test whether:

- Pharmaceutical budget impact estimates significantly impact on CVZ outcomes: increasing budgetary impact is hypothesised to increase the log-odds of non-recommendation or restriction relative to recommendation
- The use of cost-effectiveness analysis introduced in 2006 plays a role in CVZ decision-making - increasing ICERs are hypothesised to be associated with an increased log odds of restriction or non-recommendation relative to recommendation
- 'me-too' technologies negatively impact on odds of recommendation versus restriction or versus non-recommendation
- Therapeutic areas differ in their impact on the odds of recommendation versus restriction or versus non-recommendation

by CVZ are highlighted in Box 6.1. Building on from the methods described in Chapter 3, this section describes the methods used to select the sample for analysis, the outcome variable and explanatory variables considered, and the statistical techniques adopted.

### *6.2.1 Sample*

The pharmaceutical technology reviews performed by CVZ formed the basis for the sample included in this analysis. The composition of the sample was determined through the following inclusion and exclusion criteria. The sample included all drug technology appraisals (as opposed to medical devices or other interventions) made during the period 2004-2009 indicated for an adult population (≥18 years). Technology appraisals were excluded from the analysis for any of the following reasons: i) they focused on a non-adult population; ii) they appraised non-pharmaceutical interventions; iii) marketing authorisation was withdrawn; or iv) the full report was not publicly available.

### *6.2.2 Outcome variable*

To address the research question, CVZ decisions were analysed through considering HTA outcomes in three categories, where the new technology can be:

- recommended for routine use
- recommended for restricted use

or

- not recommended

**Table 6.1 Classification of CVZ coverage decisions**

| **Recommended Technology** | **Restricted Technology** | **Not Recommended Technology** |
|---|---|---|
| If technology placed in:<br><br>Reimbursement lists 1A or 1B;<br>or<br><br>Expensive drug list; | If technology placed in:<br><br>List 2<br><br>or<br><br>If patient co-payment is necessary to access medication | If words 'not recommended' were indicated on the CVZ advice<br><br>and<br><br>Technology was not included on any reimbursement list |

For the CVZ, where the decision was to place the technology in the basic package ('*basis paket*'), ie. Lists 1A or 1B, without any restriction or patient co-payment, or listed in the expensive drug list (*Dure geneesmiddelen Beleidsregel*) was considered to be recommended. Where the decision was to place the technology in the *basis paket*,

but only for use in a sub-population or with a patient co-payment, this technology was considered as restricted. And the technology was considered as not recommended when it was designated as 'not recommended' and was not included on any reimbursement list (Table 6.1).

### *6.2.3 Explanatory variables*

In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, the CVZ dataset includes 39 explanatory variables. The first set of variables was related to the technology itself – the nature of the clinical evidence available, disease characteristics, whether cost-effectiveness evidence was put forward, and if so, the characteristics of that evidence. The second set of variables captured information relating to the process by which the recommendation was issued. Included in this second set were a few variables that were specifically collected for the CVZ, such as information on whether the technology is for use within an inpatient setting (*intramurale middelen*) or an outpatient setting (*extramurale middelen*). The reasons for collecting this information is that the reimbursement regulations and payers vary according to the setting in which the technologies are used. Another variable specific to the CVZ was related to whether or not the technology was included in the expensive drug list (*Dure geneesmiddelen lijst)*. Thus, information was extracted as to whether a request was made for those technologies placed on the expensive list, and this enabled the collection of data on real-life use of the technology as a condition for reimbursement. In this category-set, data also was extracted on whether a patient co-payment was attached to a particular technology or not. Finally, the third set of variables of the data extraction form captured information on the context in which the guidance was issued (healthcare system, economic and social context), and thus aimed to capture information on health policy and socio-economic characteristics of the Netherlands.

### *6.2.4 Data extraction form*

Similar to the form utilised to extract data from NICE and SMC guidance (see Chapters 4 and 5), this form contained the definitions and decision rules used when extracting the data from the assessment documents issued by CVZ, and other data sources. The data extraction form was organized into three segments, relating to the three components of analysis that are integral to this research. The method for data extraction are described

in Chapter 3. Table 6.2 provides the list of variables extracted to create the CVZ dataset, as well as the accompanying decision rules and definitions.

**Table 6.2 CVZ dataset: Included Variables, their Definition, Data Extraction Rule and Data Sources**

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **1** | Number of RCTs considered in decision | Count | The number of distinct Randomised Controlled Trials (RCTs) that provide data related to the therapeutic indication under evaluation<br><br>Excluded: studies that are single arm, that have no randomization, or that are non-interventional. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **2** | Size of population included in RCTs | Numeric | Mean number of patients per RCT. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **3** | Length/extent of follow-up in RCT | Numeric | Mean number of weeks that data is collected on patients that entered the RCTs (see variable no. 1). | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **4** | Statistically Significant results | Categorical (yes/no/inconsistent) | Presence of statistically significant superiority of technology vs. comparator for primary endpoint(s).<br><br>If more than one RCT was considered, and the technology showed statistically significant superiority in one trial, but not in another, the results were considered to be 'inconsistent' and classified as such. RCTs designed as 'non-inferiority' studies were classified as not showing any statistically significant superiority (i.e. 'no'). | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **5** | Relevance of RCT to payer decision | Numeric | Percentage of RCTs where active comparator was used. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **6** | Number of observational studies considered in guidance | Count | Number of observational studies providing information to support study drug.<br><br>Observational studies in this circumstance are defined as studies that are non interventional (i.e. do not explicitly request the patient to take particular medication or the physician to follow particular protocol). | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **7** | Priority disease area | Categorical – yes/no | This variable aims to capture the health policy context in which payer decision is made, by capturing whether the pharmaceutical in question is linked to a disease area that is prioritized by the ministry of health. Priority disease areas were identified by examining government plans/health documents that highlight national health care system | Ministerie van Volksgezondheid, Welzijn en Sport (2003, 2007) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | | | focus. | |
| **8** | Orphan Status | Categorical – yes/no | This variable captured information on whether or not the technology was recognized by the European Medicines Agency (EMA) as an orphan designated medicine. | European Medicines Agency (2010) |
| **9** | Therapeutic Area | Categorical – 13 categories | The British National Formulary (BNF) categories were used to classify each technology into the corresponding therapeutic area. | British National Formulary (2010) |
| **10** | Prevalence of disease/clinical condition | Numeric | Reported number of patients eligible for treatment, as per the Summary Product Characteristics and indication of the medication under evaluation is indicated. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **11** | Availability of alternative therapies in current treatment setting. | Categorical – yes/no | An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **12** | Consideration of Cost Utility Analysis in guidance | Categorical – CUA performed or no CUA | Presence or absence of a cost-utility analysis. | Farmaco-economisch rapport, Vraagstelling doelmatigheidstoets |
| **13** | Incremental Cost-utility ratio of technology vs. comparator in base case | Numeric | ICER (Cost per QALY) reported in the report for base case as accepted by the CFH. This is defined as the ICER that is related to the recommendation. If more than one ICER is presented as the recommendation covers more than one population, than the ICER pertaining to the larger of the populations was reported.<br><br>If technology is reported as dominant or dominated, record as such in data extraction sheet. | Farmaco-economisch rapport, Vraagstelling doelmatigheidstoets |
| **14** | Multiple CUA/CEA models reported | Categorical - Yes/No | Whether more than one cost-utility or cost-effectiveness model was considered during the appraisal | Farmaco-economisch rapport, Vraagstelling doelmatigheidstoets |
| **15** | | If yes – provide range | If yes, report range of base case ICERs presented between the different models reported. The difference between the lowest and highest ICER will be calculated. | Farmaco-economisch rapport, Vraagstelling doelmatigheidstoets |
| **16** | Uncertainty around the base case ICER reported in submission (probabilistic) | Numeric | This should be reported as the percentage probability of acceptance at the threshold used by the agency. For the CFH the probability of the medication being cost-effective was reported at a EUR €50,000 threshold. | Farmaco-economisch rapport, Vraagstelling doelmatigheidstoets |
| **17** | Uncertainty around base case | Numeric | This should be reported as the range of ICERs (min-max) resulting | Farmaco-economisch rapport, |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | ICER reported in submission (univariate) | | from univariate sensitivity on the base case. | Vraagstelling doelmatigheidstoets |
| **18** | Non-cost per QALY cost-effectiveness analyses submitted | Yes/No | Indicates if non-cost per QALY economic analyses were submitted and reviewed. | Farmaco-economisch rapport, Vraagstelling doelmatigheidstoets |
| **19** | Anticipated budgetary impact of introduction of new technology in health care system | Numeric | Estimated annual budgetary impact of introducing new medication into the current treatment setting, if the pharmaceutical were to be introduced without any restriction. Drug cost only (per year). | Kostenprognose Rapport |
| **20** | Prevalence of disease/clinical condition | Numeric | Reported number of patients eligible for treatment, as per the Summary Product Characteristics and indication of the medication under evaluation is indicated. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f; or Kostenprognose Rapport, section 2.2 |
| **21** | Availability of alternative therapies in current treatment setting. | Categorical – yes/no | An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. | CFH-rapport , section 2a; Farmacotherapeutisch rapport, section 3, 4a-f |
| **22** | Inclusion of patient submission | Categorical – yes/no | A patient submission was considered to have been included as part of the appraisal process if a submission from a patient group was posted on the webpage pertaining to the guidance. | CVZ (2010a) |
| **23** | Number of decision makers accountable | Numeric | Captures the number of decision-makers accountable for guidance issued, as reported. | CFH (2011) |
| **24** | Cost-effectiveness evaluation component in process | Categorical – yes/no | Captures whether or not cost-effectiveness is a component of the decision-making process. If cost-effectiveness analysis is a formal part of the appraisal process, this variable was marked as 'yes'. | CFH (2006) |
| **25** | Budget impact as a component of decision-making process | Categorical – yes/no | Captures whether budget impact considerations are part of decision-making process | CFH (2006) |
| **26** | Pricing known during appraisal process | Categorical – yes/no | Captures whether the price of the technology under appraisal was known during the assessment. | CVZ (2010a) |
| **27** | Number of technologies appraised simultaneously | Count | This variable captures the number of technologies appraised simultaneously in the appraisal. | CFH-rapport , section 2a; CFH (2006) |
| **28** | Accountability of drug budget | Categorical – yes/no | The HTA agency was examined to assess whether or not the agency making the funding decisions is also accountable for the drug budget or | Sorenson et al. (2008) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | | | not. | |
| **29** | Independence of decision-making agency | Categorical – yes/no | This pertains to whether the HTA body is independent of Ministry of Health or part of it. | Sorenson et al. (2008) |
| **30** | Date guidance was issued | Numeric | Year when coverage decision was issued | Letter from the minister van Volksgezondheid, Welzijn en Sport |
| **31** | Population size – Agency coverage | Numeric | Estimate of population size within remit of the agency performing the evaluation. | Centraal Bureau voor de Statistiek (CBS)Statline |
| **32** | GDP-healthcare expenditure | Numeric (%) | Percentage of GDP spent on healthcare, during year of decision | from OECD Health Data 2009, OECD 2010 |
| **33** | Healthcare expenditure on pharmaceuticals per patient per year | Numeric (€) | Healthcare budget spent on pharmaceuticals per patient per year, during the same year in which the appraisal was published. | Centraal Bureau voor de Statistiek (CBS)Statline<br>Genees- en hulpmiddelen Informatie Project (GIP), accessed February 2010 |
| **34** | Drug funding process within healthcare system – whether centralized or decentralized | Categorical – centralized, decentralized | States whether drug funding process within the healthcare system is centralized at a national level or whether funding decisions are decentralized to the regional level | Sorenson et al. (2008) |
| **35** | Election year at time of decision | Categorical – yes/no | This variable captures whether the payer decision was made within an election year. An election year was defined as a year in which either national government or regional elections took place. | Todosijevic et al. (2010) |
| **36** | Inpatient Use | Categorical – yes/ no | This variable was extracted for the CFH and provides information on whether or not the technology was specified for use within an inpatient setting or not. | Letter from the minister van Volksgezondheid, Welzijn en Sport |
| **37** | Post-approval Study request | Categorical – yes/ no | This variable was extracted for the CFH, and provides information on whether reimbursement was granted with the condition that real-life observational data on the technology would be provided within a specified time period. | Letter from the minister van Volksgezondheid, Welzijn en Sport; |
| **38** | Co-payment | Categorical – yes/ no | This variable was extracted for the CFH to identify those technologies where patients are requested to pay a percentage of the drug cost in order to access the technology. | Medicijnkosten (2009) |
| **39** | Expensive Drug | Categorical – yes/ no | Collected for the CFH. A technology was considered to be an expensive drug if it was reported on the "Duuregeneesmiddel lijst" published by the CFH. | BELEIDSREGEL CI-891 (2006), BELEIDSREGEL CI-891 (2008) |

### *6.2.5 Statistics*

The methods for the descriptive statistics and multivariate analyses were described in Chapter 3. Descriptive statistics were calculated for each extracted variable, stratified by outcome group (recommended, restricted or not recommended). Following a descriptive analysis of the dataset, a multinomial logit regression was modelled. The objective of this analysis was to obtain a parsimonious model that best reflected the main drivers of CVZ decision-making.

## 6.3 Results

### *6.3.1 Sample characteristics*

A total of 277 drug reviews issued between January 2004 and June 2009 were retrieved from the CVZ website. Of these, 244 full submissions, representing 256 coverage decisions, were included for analysis. 33 drug reviews were excluded from the analysis for the following reasons: i) full guidance was not available (n=13); or ii) they focused on a non-adult population (n=20). In the CVZ sample, the variables with the highest proportion of incomplete entries were those related to the prevalence and budget impact estimate for the technology. Approximately 100 of 256 coverage decisions did not report prevalence and budget impact estimates for the technology.

The most common coverage decision by the CVZ was to recommend new technologies (51%), followed by restriction of funding (33%), while 16% of coverage decisions advocated not funding the technology (Table 6.3).

**Table 6.3 Outcome of CVZ Guidance issued between 2004-June 2009**

| CVZ guidance | Number of coverage decisions | Percentage |
|---|---|---|
| Recommended | 130 | 51% |
| Restricted | 86 | 33% |
| Not recommended | 40 | 16% |
| Total | 256 | 100% |

### *6.3.2 Descriptive statistics*

Clinical Evidence

Descriptive statistics for the CVZ sample are summarized in Table 6.4. On average CVZ took into consideration 3 RCTs in their review process, with an average sample size of 830 patients, and duration of 39 weeks. There appeared to be a statistically

significant difference between the outcome groups with regard to these clinical variables ($p<0.05$).

Aside from the size and number and duration of RCTs, the results of the RCTs were also captured. Of the non recommended technologies, 25% of them had demonstrated superiority in their primary endpoint, compared with 45% of recommended interventions and 35% of restricted interventions ($p= 0.046$).

The comparator used within the clinical trial programme was assessed – in particular, the percentage of comparisons made to 'active' comparators as opposed to placebo was recorded. The interventions recommended for use had a higher number of trials with active comparators (51%) than interventions that were restricted or not recommended (45%, 21% of RCTs with active comparators, respectively). The differences observed between the groups were statistically significant.

Consideration of non-randomised observational data was also recorded. Overall, in the majority of technology reviews, observational data was not considered and there were no significant differences between outcome groups.

Disease Characteristics

The CVZ review process requires information on the number of patients eligible for treatment with the technology under review. This ranged from 61,816 patients for recommended interventions to 41,087 and 268,145 patients in the restricted and not recommended interventions ($p<0.05$). The availability of alternative therapies was assessed to ascertain if it differed between recommended and restricted or not recommended interventions. In the majority of technologies appraised, an alternative was available (in 79% of cases, all decisions considered). There were minor differences between groups which were not statistically significant ($p= 0.276$).

For each technology review, in order to assess whether the type of disease impacted on decision making, the disease was coded using BNF categories – 13 categories are used ranging from cardiovascular, central nervous system to infections, and skin diseases. Across all decisions – the majority of decisions were linked to technologies for malignant disease and immunosuppression, central nervous system, and muscoskeletal and joint diseases. This differed, however, between the decision categories: for

example, musculoskeletal and joint diseases appeared to be more likely to fall in the 'restricted' category (16%) than in the recommended (6%) or not recommended groups (6%). The most common disease category in the recommended group was therapies linked to malignant disease and immunosuppression (30% of recommendations made for this disease category). However, it was also the most common disease area in the 'not recommended' group. The restricted group had a more even representation of several disease areas – malignant disease and immunosuppression (18%), muscoskeletal and joint diseases (16%), cardiovascular system (13%) and central nervous system (13%).

**Table 6.4 CVZ Coverage Decisions: Descriptive statistics for extracted variables, by coverage decision (recommended, restricted, not recommended)**

| Variable | Total (n=256) | | | Recommended (n=130) | | | Restricted (N=86) | | | Not Recommended (n=40) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | |
| Number of RCTs considered in decision | 2.6 | 2.3 | 3.0 | 2.2 | 1.8 | 2.5 | 2.8 | 2.2 | 3.5 | 3.8 | 2.0 | 5.6 |
| Size of population included in RCTs | 830 | 494 | 1165 | 525 | 366 | 684 | 1364 | 447 | 2281 | 588 | 232 | 945 |
| Length/extent of follow-up in RCT (weeks) | 39 | 33 | 46 | 44 | 34 | 54 | 31 | 24 | 38 | 44 | 24 | 65 |
| Statistically Significant results - yes | 39% | 33% | 45% | 45% | 37% | 54% | 35% | 25% | 45% | 25% | 11% | 39% |
| *No* | 29% | 23% | 34% | 20% | 13% | 27% | 35% | 25% | 45% | 45% | 29% | 61% |
| *inconsistent* | 17% | 13% | 22% | 18% | 11% | 24% | 17% | 9% | 26% | 15% | 3% | 27% |
| Use of Active Comparator in RCT | 44% | 38% | 51% | 51% | 42% | 60% | 45% | 35% | 55% | 21% | 8% | 34% |
| Number of observational studies considered in guidance | 0.6 | 0.4 | 0.7 | 0.6 | 0.3 | 0.8 | 0.6 | 0.2 | 0.9 | 0.5 | 0.1 | 0.8 |
| Consideration of Cost Utility Analysis in guidance | 11% | 7% | 15% | 12% | 7% | 18% | 7% | 1% | 12% | 18% | 5% | 30% |
| Incremental Cost-effectiveness ratio of technology vs. | € 36,621 | € 10,217 | € 63,024 | € 51,854 | € 2,005 | € 101,704 | € 17,963 | € 11,339 | € 24,588 | € 18,502 | € 9,103 | € 27,901 |

| Variable | Total (n=256) | | | Recommended (n=130) | | | Restricted (N=86) | | | Not Recommended (n=40) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | |
| comparator in base case | | | | | | | | | | | | |
| More than one CUA submitted | 1% | 0% | 2% | 2% | 0% | 5% | 0% | 0% | 0% | 0% | 0% | 0% |
| If More than one CUA submitted - low range | € 100,592 | -€ 234,954 | € 436,138 | € 100,592 | -€ 234,954 | € 436,138 | . | . | . | . | . | . |
| If More than one CUA submitted - high range | € 261,851 | € 238,338 | € 285,363 | € 261,851 | € 238,338 | € 285,363 | . | . | . | . | . | . |
| Uncertainty around the base case ICER reported in submission (probabilistic) | 66% | 42% | 90% | 79% | 43% | 115% | 52% | -66% | 170% | 55% | -168% | 277% |
| Uncertainty around base case ICER reported in submission (univariate) Low | € 17,954 | € 5,128 | € 30,779 | € 31,462 | -€ 19,406 | € 82,331 | € 13,402 | -€ 16,808 | € 43,611 | € 10,465 | € 2,373 | € 18,557 |
| Uncertainty around base case ICER reported in submission (univariate) High | € 109,737 | -€ 7,514 | € 226,987 | € 192,709 | -€ 106,725 | € 492,142 | € 48,132 | -€ 43,192 | € 139,455 | € 43,739 | -€ 539 | € 88,017 |
| Prevalence of disease/clinical condition | 94,543 | 31,394 | 157,693 | 61,816 | - 16,081 | 139,713 | 41,087 | 8,932 | 73,242 | 268,145 | 11,852 | 524,440 |
| Potential budgetary impact (million) | € 36.4 | € 6.9 | € 65.9 | € 6.7 | € 4.3 | € 9.0 | €66.0 | -€ 17.6 | €76.0 | € 268,145 | -€ 26.5 | € 175 |
| Societal | 1% | 0% | 1% | 1% | -1% | 3% | 0% | 0% | 0% | 0% | 0% | 0% |

| Variable | Total (n=256) | | | Recommended (n=130) | | | Restricted (N=86) | | | Not Recommended (n=40) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | |
| Perspective adopted | | | | | | | | | | | | |
| Availability of alternative therapies in current treatment setting. | 79% | 74% | 84% | 76% | 68% | 83% | 85% | 77% | 93% | 78% | 64% | 91% |
| Inclusion of patient submission | 4% | 2% | 6% | 2% | 0% | 5% | 2% | -1% | 6% | 13% | 2% | 23% |
| Number of Decision Makers Accountable | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| Cost-effectiveness evaluation component in process | 67% | 61% | 73% | 65% | 57% | 74% | 64% | 54% | 74% | 78% | 64% | 91% |
| Budget impact as a component of decision-making process | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Price of technology known during appraisal | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Number of technologys appraised in same appraisal | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

| Variable | Total (n=256) | | | Recommended (n=130) | | | Restricted (N=86) | | | Not Recommended (n=40) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Date guidance was issued | 2006 | 2006 | 2006 | 2006 | 2006 | 2006 | 2006 | 2006 | 2006 | 2007 | 2006 | 2007 |
| Population size – Agency coverage (millions) | 16.30 | 16.30 | 16.40 | 16.30 | 16.30 | 16.40 | 16.30 | 16.30 | 16.30 | 16.40 | 16.30 | 16.40 |
| GDP-healthcare expenditure | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% |
| Healthcare expenditure on pharmaceuticals | € 294 | € 291 | € 296 | € 293 | € 289 | € 296 | € 292 | € 287 | € 296 | € 302 | € 296 | € 309 |
| Election year at time of decision | 20% | 15% | 24% | 23% | 16% | 30% | 20% | 11% | 28% | 8% | -1% | 16% |
| Priority disease area | 55% | 49% | 61% | 61% | 52% | 69% | 43% | 32% | 54% | 60% | 44% | 76% |
| Orphan Designated | 9% | 5% | 12% | 11% | 5% | 16% | 9% | 3% | 16% | 0% | 0% | 0% |
| Technology has EU Marketing Authorisation | 92% | 88% | 95% | 92% | 88% | 97% | 94% | 89% | 99% | 85% | 73% | 97% |
| Future Cost-Effectiveness Analyses requested | 9% | 5% | 13% | 16% | 10% | 23% | 1% | -1% | 3% | 3% | -3% | 8% |
| Expensive Drug | 12% | 8% | 16% | 23% | 16% | 30% | 0% | 0% | 0% | 0% | 0% | 0% |
| Patient Copayment | 4% | 2% | 7% | 2% | -1% | 4% | 9% | 3% | 16% | 3% | -3% | 8% |

| Variable | Total (n=256) | | | Recommended (n=130) | | | Restricted (N=86) | | | Not Recommended (n=40) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | | mean | 95% Confidence interval | |
| needed | | | | | | | | | | | | |
| BNF1 cardiovascular system | 9% | 5% | 12% | 4% | 0% | 7% | 15% | 7% | 23% | 10% | 0% | 20% |
| BNF2 central nervous system | 16% | 12% | 21% | 15% | 9% | 22% | 15% | 7% | 23% | 23% | 9% | 36% |
| BNF3 ear, nose and oropharynx | 1% | 0% | 2% | 2% | 0% | 5% | 0% | 0% | 0% | 0% | 0% | 0% |
| BNF4 endocrine system | 6% | 3% | 9% | 5% | 1% | 8% | 8% | 2% | 14% | 5% | -2% | 12% |
| BNF5 eye | 2% | 0% | 3% | 3% | 0% | 6% | 0% | 0% | 0% | 0% | 0% | 0% |
| BNF6 gastro-intestinal system | 5% | 2% | 7% | 5% | 1% | 9% | 5% | 0% | 9% | 3% | -3% | 8% |
| BNF7 infections | 9% | 5% | 12% | 6% | 2% | 10% | 12% | 5% | 19% | 10% | 0% | 20% |
| BNF9 musculoskeletal and joint diseases | 10% | 6% | 14% | 7% | 3% | 11% | 19% | 10% | 27% | 3% | -3% | 8% |
| BNF10 nutrition and blood | 7% | 4% | 10% | 8% | 3% | 12% | 8% | 2% | 14% | 3% | -3% | 8% |
| BNF11 obstetrics, gynaecology, and urinary-tract disorders | 3% | 1% | 5% | 1% | -1% | 2% | 7% | 1% | 12% | 3% | -3% | 8% |
| BNF12 respiratory system | 4% | 1% | 6% | 2% | -1% | 4% | 0% | 0% | 0% | 18% | 5% | 30% |
| BNF13 skin | 5% | 2% | 7% | 5% | 1% | 9% | 2% | -1% | 6% | 8% | -1% | 16% |

Orphan designation is granted by the EMA to technologies that fulfil specific criteria, namely related to the prevalence of the disease, the clinical profile of the disease and the level of unmet need. Across 2004-2009, 9% of technologies reviewed by the CVZ had an orphan designation. A higher proportion of orphan designated therapies were in the recommended and restricted groups (11% and 9% respectively), compared to the not recommended interventions (0%) ($p = 0.10$).

Economic Evidence

A range of economic related variables were included for analysis. Only 11% of CVZ decisions were backed by the use of a CUA. There was no statistically significant difference between recommended (12%), restricted (7%) or not recommended interventions (18%). Only 1% of reviews considered more than one CUA – in almost all cases a single CUA was reviewed. For the interventions supported by a CUA, the average incremental cost-effectiveness ratio (ICER) across all groups was €36,620. The mean ICER differed between the interventions, but not in the anticipated direction. The mean ICER for the recommended interventions was €51,854, compared to the restricted interventions (mean ICER of €37,481), and the interventions not recommended for use (mean ICER of €18,501). However, these differences were not statistically significant.

For those interventions that reported an ICER, information on the uncertainty around base-case ICER estimates was collected. This was done by recording results of probabilistic sensitivity analyses (the probability of ICER remaining below a set threshold), and by recording univariate sensitivity analyses (lowest ICER, highest ICER). The probability of the ICER remaining below €20,000 was 79% for recommended interventions, 52% for restricted interventions and 55% for interventions not recommended for use. However, these differences were not statistically significant across any of the tests performed, and the sample size was very small (n=14). With regard to univariate sensitivity analyses, the statistical tests performed do not suggest a statistically significant difference between the decision groups. Aside from examining the results of CUA considered by the CVZ review committee, information was also captured on whether alternative cost-effectiveness models (non-CUA) were considered in the decision-making process; 15% of interventions were supported by non-CUA models. The use of non-CUA analyses occurred in 15% of recommended interventions, 9% of restricted interventions and 28% of interventions not recommended

for use. The differences observed between decision outcome groups were statistically significant (p=0.016).

The potential budget impact of a positive recommendation also was analysed - the mean estimated maximum yearly budget impact across all decisions was in the order of € 36 million. This ranged from €7 million for the recommended interventions to €66 and €75 million for the restricted and not recommended interventions, respectively (p=0.0001).

CVZ process characteristics

A series of variables related to the decision-making process were recorded. Less than 1% of decisions formally considered a societal perspective in the review process. In very few decisions (4%), patient-group submissions were formally considered in the appraisal. The CVZ committee operates on a pre-defined number of 20 committee members who can serve on the committee for eight years. It was not clear from the information provided on the webpage, how the committee processed the technology reviews (i.e. if there were lead reviewers etc). In all of the decisions there was a budget impact component. The cost-effectiveness component was added as part of the review process in 2006, and is applicable in those instances where the manufacturer requests reimbursement for a compound in the 1B reimbursement list, or listing on the expensive drug list. The CVZ review process considers only one compound at a time (one exception noted in one appraisal on a group of anti-allergen therapies). None of the above variables differed significantly between outcome groups.

Aside from reviewing therapies that have received marketing authorisation, the CVZ also has the role of reviewing technologies that do not have marketing authorisation, but that could be of value as 'last resort' treatment options for patients with very severe illnesses resistant to standard therapy. This is the case in 8% of CVZ decisions reviewed in this research project. This percentage was constant across the three groups.

In the Netherlands, patients can be asked to provide a co-payment for their prescription. This is determined by whether or not a certain technology falls within a 'basis pakket' of insurance or not. For the purposes of this research, it was considered that, for those technologies where a patient co-payment is necessary, these would fall under the 'restricted' category. The rationale for this coding is that although the medication is

recommended for use by the Ministry of Health, it is not reimbursed 100% and therefore there is a restriction on its access, which is in fact the co-payment. Of the 256 decisions made, only 4% of these are associated with a patient co-payment. These were all placed within the 'restricted' decision category.

Socio-economic context of CVZ decision-making

A series of variables were recorded to capture the socio-economic context in which CVZ decisions were made. In particular, information was recorded on the year of the appraisal (mean 2006), size of population under the CVZ's remit (approximately 16.3 million), % of GDP spent on healthcare (10% on average across groups), whether appraisal coincided with an election year and whether the technology under appraisal was linked to a disease considered to be a 'priority' by the Ministry of Health (55% of technologies appraised were directly linked to a priority disease area). There was strong correlation between the year of appraisal and the other socio-economic factors. The year of appraisal, population size and healthcare expenditure on pharmaceuticals, appeared to vary significantly between outcome groups ($p=0.05$). The rest of the variables did not differ significantly between outcome groups.

Summary of descriptive statistics

In total, within the CVZ sample, of the 40 explanatory variables that were explored, descriptive analysis suggests that 16 factors may play an important contributing factor in determining SMC decision-making (Table 6.5). For these variables, statistically significant differences were observed between interventions that were recommended, restricted and not recommended ($p \leq 0.05$). Highly statistically significant differences ($p \leq 0.01$) were observed for variables highlighted in bold in the Table 6.5.

**Table 6.5 CVZ Descriptive statistics: statistically significant variables (p≤ 0.05)**

| Evidence factors | Variables |
|---|---|
| Clinical Package | Number of RCT<br>Size of RCT<br>RCT duration<br>**Superiority demonstrated**<br>Use of active comparator |
| Economic Package | Non-CUA analyses submitted<br>**Budget Impact** |
| Disease characteristics | **Prevalence of disease**<br>Disease categories (BNF): musculoskeletal and joint diseases; obstetrics, gynaecology and urinary-tract disorders; respiratory system; cancer; cardiovascular system |
| **Process factors** | **Presence of patient submission**<br>Priority disease<br>Request for future CEA<br>Patient co-payment |
| **Socio-economic context factors** | Date of Review<br>Pharmaceutical expenditure per patient<br>Size of the national population |

Note: variables in bold text were statistically significant at the $p \leq 0.01$ level

### *6.3.3 Multivariate analysis*

Following the model specification process described in Chapter 3 which included the development of a preliminary model (Appendix C), the CVZ base case model was developed which contains nine variables, including a mixture of clinical, economic and process variables (Table 6.6). The model resulted in a pseudo R-squared of 0.17, suggesting that the model explains approximately 17% of the variability in CVZ coverage decisions. Clinical variables that had a significant impact on the CVZ model include the presence of an active comparator, which decreased the log odds of a restriction or non-recommendation relative to recommendation, although this effect was only statistically significant for the latter ($p=0.239$and $p<0.001$ respectively). Demonstration of superiority in the clinical trial also decreased the odds of a restriction or non-recommendation relative to recommendation ($p=0.022$, $p<0.001$ respectively). The lack of information on the duration of the RCT increased the odds of a non-recommendation ($p=0.001$) and restriction (not statistically significant) relative to recommendation. The budget impact of the technology appeared to impact significantly on CVZ decision-making: a unit increase in the budget impact increased the probability of a restriction ($p=0.051$) relative to recommendation, but was not statistically significant on the log odds of a non recommendation relative to recommendation. Technologies indicated for the treatment of cancer decreased the log odds of a restriction or non recommendation relative to recommendation, and this impact was statistically significant on the log odds of a restriction ($p<0.001$) and borderline

significant on the log odds of a non recommendation relative to recommendation (p=0.177). Technologies indicated for the treatment of cardiovascular disease, and obstetrics/gynaecology/ urinary-tract disorders increased the probability of a restriction relative to recommendation, and this was statistically significant.

**Table 6.6 Multivariate analysis of CVZ coverage decisions 2004-2009: base case model results**

| **Restricted** | **Log Odds** | **P value** | **95% Confidence Interval** | |
|---|---|---|---|---|
| Use of active comparator in RCT | -0.49 | 0.239 | -1.31 | 0.33 |
| Demonstrated clinical superiority in RCT | -0.82 | 0.022 | -1.53 | -0.12 |
| Budgetary Impact | 0.0069 | 0.051 | -0.000031 | 0.014 |
| Cancer therapy | -1.59 | <0.001 | -2.47 | -0.72 |
| Therapies for cardiovascular diseases | 1.40 | 0.017 | 0.25 | 2.55 |
| Therapies for obstetrics, gynaecology, and urinary-tract disorders | 2.40 | 0.032 | 0.20 | 4.59 |
| Prevalence of target population | -0.0000014 | 0.091 | -0.0000029 | 0.0000002 |
| Patient Submission | -0.15 | 0.879 | -2.08 | 1.78 |
| Lack of data on duration of RCT | 0.24 | 0.560 | -0.58 | 1.06 |
| Constant | 0.28 | 0.459 | -0.46 | 1.03 |
| **Not Recommended** | **Log Odds** | **P value** | **95% Confidence Interval** | |
| Use of active comparator in RCT | -2.54 | <0.001 | -3.84 | -1.24 |
| Demonstrated clinical superiority in RCT | -1.85 | <0.001 | -2.82 | -0.89 |
| Budgetary Impact | 0.0047 | 0.217 | -0.0028 | 0.012 |
| Cancer therapy | -0.70 | 0.177 | -1.71 | 0.31 |
| Therapies for cardiovascular diseases | 0.82 | 0.338 | -0.86 | 2.51 |
| Therapies for obstetrics, gynaecology, and urinary-tract disorders | 2.20 | 0.141 | -0.73 | 5.12 |
| Prevalence of target population | -0.000000016 | 0.982 | -0.0000014 | 0.0000014 |
| Patient Submission | 1.79 | 0.032 | 0.16 | 3.41 |
| Lack of data on duration of RCT | 1.78 | 0.001 | 0.74 | 2.81 |
| Constant | 0.066 | 0.878 | -0.78 | 0.91 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.17.

Impact of alternative model specifications – sensitivity analyses

Sensitivity analyses were conducted on the base-case model of CVZ coverage decisions. This included i) examining the impact of a binary rather than three-category outcome variable; ii) restricting the base case analysis to complete observations, thus excluding observations with imputed values, and (iii) examining the impact of assuming the coverage decisions are ordinal.

In the first sensitivity analysis, the model was run using a binary outcome variable. This was done by considering a 'covered' and 'not covered' approach: the recommended and restricted categories were grouped together, and the not recommended category was kept as in the base case analysis. A logistic regression was performed examining the log likelihood and the odds of coverage versus no coverage by the healthcare system (Table 6.7).

In this binary model, the use of an active comparator in the clinical trial and the demonstration of superiority remained significant variables, as did whether a patient submission was completed. The budget impact, duration of the RCT and whether the technology was indicated for cancer, cardiovascular disease or obstetrics/gynaecology

diseases were no longer significant variables. The pseudo R-squared was 0.21; suggesting that the model can explain approximately 21% of the variability in CVZ coverage decisions, higher than the pseudo R-squared observed in the base-case model (0.17). New variables not present in the base case model that showed significance in this sensitivity analyses were: date of appraisal, and therapies for infectious diseases. The use of an active comparator in the clinical trials and the demonstration of superiority in the primary endpoint increased the probability of listing vs. non-listing (p=0.003 and p=0.058 respectively). The effect of these explanatory variables observed in this sensitivity analysis was similar to that observed in the base-case analysis. The date of appraisal, the inclusion of a patient submission and if the technology was indicated for the treatment of infectious diseases decreased the probability of coverage (p=0.081, p=0.017, p=0.002 respectively). In the base-case model, the impact of patient submissions differed on the log odds of restriction or non recommendation. The presence of patient submissions increased the log odds of a recommendation relative to a restriction (although this was not statistically significant), while the presence of submissions increased the log odds of non-recommendation (p=0.032). Using a binary outcome category removes the ability to explore the impact of the same factor on the odds of different types of coverage decisions.

**Table 6.7 Sensitivity Analysis 1. Multivariate analysis of CVZ coverage decisions 2004-2009: sensivitiy analysis using binary outcome variable**

| | Odds of Listing | P value | 95% Confidence Interval | |
|---|---|---|---|---|
| Use of active comparator in RCT | 1.70 | 0.003 | 0.58 | 2.82 |
| Demonstrated clinical superiority in RCT | 0.80 | 0.058 | -0.027 | 1.63 |
| Therapy for Infectious diseases | -2.84 | 0.002 | -4.60 | -1.07 |
| Prevalence of target population | -0.00000077 | 0.11 | -0.0000017 | 0.00000018 |
| Patient Submission | -1.74 | 0.017 | -3.16 | -0.31 |
| Date of Appraisal | -0.24 | 0.081 | -0.50 | 0.03 |
| Constant | 477.57 | 0.081 | -58.42 | 1013.56 |

In the second sensitivity analysis, the observations with missing entries were removed from the sample. This reduced the sample of observations for analysis to 98 appraisals (vs. 256 in the base case sample) (Table 6.9). The results suggest that, firstly, the majority of observations within the CVZ have unavailable data for at least one of the explanatory variables. In addition, the results suggest that the model's ability to explain the variability in CVZ decision-making is reduced, as the pseudo R-squared is 0.12, lower than the base case (0.17). A unit increase in the budget impact has a border-line statistically significant effect on the log odds of a restriction and non-recommendation. This echoes the results of the base-case model, although the statistical significance of

the budget impact estimate is weaker in this sensitivity analysis. The demonstration of superiority or the duration of the RCT maintains a similar direction of effect and statistical significance as observed in the base-case model. The use of an active comparator continues to be important in increasing the log odds of a recommendation vis a vis a restriction or non recommendation, although not statistically significant on the log odds of restriction. Similarly to the base case, when the technology is indicated for cardiovascular diseases it increases the log odds of a restriction (p=0.03).

**Table 6.8 Sensitivity Analysis 2. Multivariate analysis of CVZ coverage decisions 2004-2009: sensivitiy analysis excluding incomplete observations from sample of analysis (n=98)**

| **Restricted** | **Log Odds** | **P value** | **95% Confidence Interval** | |
|---|---|---|---|---|
| Use of active comparator in RCT | -0.68 | 0.292 | -1.96 | 0.59 |
| Demonstrated clinical superiority in RCT | -0.39 | 0.480 | -1.45 | 0.68 |
| Budgetary Impact | 0.03 | 0.136 | -0.0086 | 0.063 |
| Therapies for cardiovascular diseases | 1.57 | 0.030 | 0.15 | 3.00 |
| Constant | 382.89 | 0.273 | -302.11 | 1067.89 |
| **Not Recommended** | **Log Odds** | **P value** | **95% Confidence Interval** | |
| Use of active comparator in RCT | -1.68 | 0.044 | -3.31 | -0.04 |
| Demonstrated clinical superiority in RCT | -1.34 | 0.027 | -2.54 | -0.15 |
| Budgetary Impact | 0.03 | 0.139 | -0.0088 | 0.063 |
| Therapies for cardiovascular diseases | 1.11 | 0.208 | -0.62 | 2.83 |
| Constant | -252.23 | 0.528 | -1035.63 | 531.18 |

Note: Recommended technologies are the reference case

In the third sensitivity analysis, ordinality of the outcome variable was assumed. This is in contrast with the base case analysis, where ordinality was not assumed and multinomial logistic regression was used. In this sensitivity analysis ordinal logistic regression was used. The detailed results of this analysis are provided in Appendix G, and show that while they yielded a slightly higher pseudo R-squared than in the base case model (0.19 vs. 0.17), the explanatory variables remained significant as observed in the base case model of CVZ decision-making.

## 6.4 Discussion

The overall objective of this chapter was to examine the factors that influence decisions made by CVZ to recommend, restrict or not recommend new technologies for use in the Dutch healthcare system. In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, and in light of evidence review presented in Chapter 2, a wide range of explanatory variables were included in the analysis, reflecting the clinical and economic characteristics of the technology under appraisal, the appraisal process itself and the socio-economic context in which the CVZ operates. In addition to

the general aims of the research, specific hypotheses relevant for CVZ decision-making were explored and are discussed below.

The analysis of CVZ included 256 pharmaceutical technology appraisals. The internal validity of the results obtained in this analysis was examined in two ways - firstly, by comparing the results with published analyses of CVZ decision-making, and secondly by sharing the base-case model results for review with members of the CVZ (Dr. Graaff, M; Dr. Goettsch, W, Dr. S. Kleijnen,)[1]. The aim of this interaction was to ascertain if the CVZ characteristics were accurately captured in the sample used for the analysis, if the approach to the analysis was clear and in particular, to gauge the reaction to the model results and potential for suggestions or additional analyses.

### *6.4.1 Pattern of CVZ coverage decisions*

The most common coverage decision by the CVZ was to recommend new technologies (51%), followed by restriction of funding (33%), while 16% of coverage decisions advocated for not funding the technology. No published information on CVZ decision-making was identified to allow for a comparison of the pattern of coverage decision observed in this analysis.

### *6.4.2 Role of clinical evidence and disease characteristics on CVZ decision-making*

The use of an active comparator and demonstration of clinical superiority were important factors that significantly impacted on CVZ decision-making. The majority (51%) of recommended technologies were supported by RCTs with active comparator arms, as opposed to 45% of restricted technologies and 21% of technologies not recommended for funding. Not only were recommended technologies more likely to be compared to active treatments, but the RCTs also demonstrate clinical superiority of the technology 45% of appraisals, compared with 35% and 25% for restricted and not recommended technologies. Other RCT characteristics (size, duration) appeared to be less important in CVZ decision-making. Of interest in the multivariate analysis is the effect of lack of evidence of trial duration which statistically significantly increases the log odds of non-recommendation relative to recommendation. This represents 37

[1] Dr. Martin van der Graaff, Secretary medicines evaluation committee (CVZ); Dr. Wim Goettsch; Sarah Kleijnen (M.Sc. ) Project coordinator EUnetHTA WP5. Interviewee: Karin Cerri. Interviewed by telephone, on January 6th 2011. Meeting minutes are provided in Appendix E.

appraisals in which information on duration of RCT was lacking. Amongst this subset of appraisals, 56% (21 of 37) appraisals supported largely by non-interventional non-randomised clinical evidence, and 46% of appraisals were focused on technologies for which there were no alternative treatments available.

Interestingly, in this model, the impact of the therapy area for which the technology was indicated played an important role in coverage decisions. It was hypothesised that differences would be observed in the outcome variable according to the therapeutic area which was targeted by the appraised technologies. In a discrete choice experiment amongst Dutch healthcare professionals, the analysis of choices made suggested that severity of disease was one of the most significant criteria driving coverage decisions (Koopmanschap et al. 2010). In the CVZ multivariate analysis presented in this chapter, cancer therapies, which could be approximated to represent severe disease, significantly decreased the log odds of restriction relative to recommendation. In contrast, technologies for the treatment of cardiovascular disease, and obstetrics/gynaecology/urinary-tract disorders increased the probability of a restriction.

### *6.4.3 Role of Economic evidence in CVZ decision-making*

It was hypothesised that the introduction of cost-effectiveness analysis criteria in CVZ decision-making in 2006 would play a role in its decision-making. However, unlike NICE and SMC models, the role of the ICER is not observed in this analysis. Only 11% of CVZ appraisals reported an ICER, and amongst those technologies for which ICERs were reported there was no statistically significant difference between outcome variables in the ICERs. However, in Koopmanschap et al. (2010), results of the discrete choice experiment suggest that CUA is an important criterion for Dutch healthcare decision-makers. Plausible explanations for this difference could be due to variation between hypothetical reimbursement decisions versus real-life decision-making, and the fact that CUA was first introduced in the CVZ process in 2005, and is only utilised as a criterion for inclusion of technologies on List 1B. It has been highlighted that the role of CUA within the CVZ appraisal process may increase in the future (Dr. Graaff, M; Dr. Goettsch, W, Dr. S. Kleijnen)[2].

[2] Dr. Martin van der Graaff, Secretary medicines evaluation committee (CVZ); Dr. Wim Goettsch; Sarah Kleijnen (M.Sc. ) Project coordinator EUnetHTA WP5. Interviewee: Karin Cerri. Interviewed by telephone, on January 6th 2011. Meeting minutes are provided in Appendix E.

An additional hypothesis was made about the role of budgetary impact evidence in CVZ decision-making. Specifically, increasing budgetary impact was hypothesised to increase the log-odds of non-recommendation or restriction relative to recommendation. The analysis showed that the size of the estimated budget impact associated with the introduction of the technology had a significant effect on CVZ decisions: a unit increase in the budget impact increased the log odds of a restriction relative to recommendation. The mean estimated budget impact for technologies recommended by the CVZ was €6.7 million compared with €66.0 million for restricted technologies and €76 million for not recommended technologies. However, the effect of budgetary impact considerations was not statistically significant on the odds of a non-recommendation relative to recommendation, suggesting there are other factors that better explain non-recommendations than budgetary impact considerations. For example, the size of the eligible patient population for restricted technologies is lower (mean eligible population is 41,087 patients) compared to non-recommended technologies (268,145). In addition, the clinical evidence for restricted technologies is supported by a higher mean number of trials (2.8 vs. 3.8 trials), with larger mean sample size (1363 vs. 588 patients) and more frequently use active comparators than non-recommended technologies (45% vs. 21%).

### *6.4.4 Role of process and socio-economic context variables in CVZ decision-making*

It is of interest to note that, with the exception of the patient submissions and the introduction of the ICER discussed above, none of the other appraisal process characteristics or socio-economic factors appeared to have significant effects on coverage decisions. This suggests the general stability of the appraisal process and socio-economic context of decision-making, and therefore the lack of significant variation within the period for which coverage decisions were extracted. In addition, it should be noted that socio-economic factors varied at the agency level, rather than at the decision-level: thus, the degree of variability was substantially reduced to annual changes in socio-economic factors. The relative importance of socio-economic and process factors will be examined further in the pooled analysis of coverage decisions from all four HTA bodies, presented in Chapter 8.

### *6.4.5 Limitations*

When examining the results of the multivariate analyses, there are several limitations that need to be taken into account. An important limitation is related to the ability to

access information that is publicly available, as the CVZ presents in the public domain the information that corresponds to the final recommendation. For instance, a technology that is recommended for GVS List 1A, meaning that it is clustered with therapies already available within the Dutch healthcare system, will have the report/information used to support this coverage decision are made publicly available. However, it is possible that a manufacturer may have submitted a request for the drug to be included in GVS List 1B i.e. where the technology is found to have therapeutic benefit to the degree that it is not clustered with existing therapies. This type of submission would have required the technology to demonstrate its cost-effectiveness and thus the manufacturer may have submitted cost-effectiveness analyses. However, if the decision by the CVZ/CVZ was that there was no added therapeutic benefit to justify inclusion in the 1B list, then the cost-effectiveness criteria would not have been applied, and therefore the information about the cost-effectiveness analyses that were performed would not be disclosed in the public domain. From one perspective, the lack of access to cost-effectiveness data that was not actually taken into account in the coverage decision does not have significant implications for the analysis, given that the aim is to identify those factors that impact on decisions. However, this perspective assumes that the availability of cost-effectiveness data has no impact on the decision to cluster or not cluster the technology. To some degree the implications of incomplete observations was addressed through imputation techniques, coupled with the use of dummy variables to identify the impact of incomplete data on coverage decisions, and an additional sensitivity analysis in which the model was restricted to those technologies with complete observations. This sensitivity analysis highlighted that the majority of observations from the CVZ were incomplete (98 of 256 were included in the analysis). This suggests that, on the one hand, the variables chosen for extraction, while standard across the various HTA bodies, may not have been appropriate for the CVZ given their reporting and decision-structure. It also suggests that there may be substantial additional considerations and data provided to the CVZ that are not disclosed (e.g. patient submissions, physician organisation interaction, cost-effectiveness analyses etc). Therefore, greater transparency in the evidence received or submitted may help to further increase the understanding of CVZ decision-making.

As with the SMC appraisal process, the CVZ also relies on manufacturer submissions in formulating its advice. There is no third party or significant additional new analysis performed on the evidence submitted by the manufacturer. Given the lack of

accessibility to manufacturer submissions in the public domain, it was not possible to take into account in the analyses to what degree the CVZ recommendation was driven by the manufacturer submission strategy relative to CVZ decision-making criteria. For example, for a technology which was accepted in the GVS 1A list (clustered technologies), it was not possible to ascertain if the inclusion in this list was proposed by the CVZ, or whether the inclusion on GVS 1A list was proposed by the manufacturer in their submission. Despite the lack of access to such information, it does not detract from the possibility of being able to assess the degree to which key characteristics and factors vary according to the coverage decision made.

While the base-case model was based on a three-category outcome variable, a sensitivity analysis was conducted to assess the effect of using a binary outcome category as has been done in published literature (Devlin and Parkin 2004; Clement et al. 2009). The results of this sensitivity analysis suggests that several of the explanatory variables which were important in the base-case model continued to be significant in this binary model. For example, the use of an active comparator and the demonstration of clinical superiority maintained their effect and significance, suggesting that the role of these variables in explaining CVZ coverage decision-making is robust to changes in model-specification. Other variables, including the budget impact, were no longer found to have a significant effect in this sensitivity analysis, suggesting that the use of binary outcome category does not allow for a more detailed exploration of the impact of the budget impact on different coverage decisions, and thus when examined in this sensitivity analysis its overall impact was not significant. Moreover, variables that were not included in the base-case model were found to have a significant role in this sensitivity analysis – this included the year of appraisal and technologies indicated for the treatment of infectious diseases. This suggests that the use of binary outcome categories can yield an alternative perspective on CVZ decision-making, at the expense of reducing visibility on the impact of explanatory variables on specific types of coverage decisions.

In summary, the overall objective of this chapter was to examine the factors that influence decisions made by CVZ to recommend, restrict or not recommend pharmaceutical technologies for use in the Dutch healthcare systems, with a focus on research hypotheses specific to CVZ decision-making. The results suggest that the variability in coverage decisions observed can be explained by a combination of

clinical, disease and economic factors. The analysis confirmed the hypothesis that pharmaceutical budget impact estimates impact significantly on CVZ outcomes – increasing budgetary impact increases the odds of restriction relative to recommendation. The differential impact of the therapeutic area for which the technology is indicated was also observed – cancer therapies decreased the log odds of restriction and non-recommendation relative to recommendation, while therapies for cardiovascular disease, for example, increased the log odds off restriction and non-recommendation relative to recommendation. The analyses did not support the hypothesis that the introduction of cost-effectiveness component in the CVZ process has impacted on its decision-making.

## 6.5 References

Centraal Bureau voor de Statistiek (CBS). 2010. Statline, Population Statistics (Bevolking, Kerncijferrs). http://statline.cbs.nl/StatWeb/ publication/? VW =T&DM =SLNL&PA=37296ned&D1=a&D2=0,10,20,30,40,50,(l-1)-l&HD=110130-2055&HDR=G1&STB=T. Viewed 26 July2010.

Genees- en hulpmiddelen Informatie Project (GIP) 2010. "GIPdatabank: Informatie over genees- en hulpmiddelen". http://www.gipdatabank.nl/ . Viewed 17 February 2010

College Tarieven Gezondsheidszorg ZorgAutoriteit. 2006. Beleidsregel CI-891, Bijlage 2 CTG-ZAio beleidsregel dure geneesmiddelen 2006.

———. 2008. Beleidsregel CI-891 Overzicht tijdelijk opgenomen intramural geneesmiddelen in de NZa beleidsregel ´Dure geneesmiddelen´ en de bijbehorende tijdstermijnen.

Commissie Farmaceutische Hulp. 2006. Farmacotherapeutisch rapport adalimumab (Humira©) bij de indicatie spondylitis ankylopoëtica. http://www.cvz.nl/binaries/live/cvzinternet/hst_content/nl/documenten/CFH-rapporten/2006/CVZ0607+adalimumab+humira.pdf

———. 2011. "Samenstelling CFH". http://www.cvz.nl/zorgpakket/CVZagenda/commissie/commissie.html. Viewed 22 January 2011.

College voor zorgverzekeringen (CVZ).2010a. Procedure beoordeling extramurale Geneesmiddelen. http://www.fk.cvz.nl/).

———.2010b. CVZ-criteria voor beoordeling therapeutische waarde.

European Medicines Agency. 2009. “Human Medicines”. http://www.ema.europa.eu/ . Viewed on 5 January 2011.

———. 2011. European public assessment reports. http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&murl=menus/medicines/medicines.jsp&mid=WC0b01ac058001d125. Viewed between 1st of June 2009-30th March 2010.

Joint Formulary Committee. 2010. British National Formulary. 60 ed. London: British Medical Association and Royal Pharmaceutical Society. http://bnf.org/bnf/index.htm

Medicijnkosten. 2009. “Medicijnkosten – een themasite van het College voor zorgverzekeringen”. http://www.medicijnkosten.nl/. Viewed 6 June 2009.

Ministerie van Volksgezondheid, Welzijn en Sport. 2003. Langer Gezond Leven – ook een kwestie van gezond gedrag 2004-2007. Publicatie Postbus 51, Rijsoverheid. http://www.rijksoverheid.nl/documenten-en-publicaties/publicaties-pb51/langer-gezond-leven-2004-2007.html

———. 2007. Kaderbrief 2007-2011 - visie op gezondheid en preventie. Publicatie Postbus 51, Rijsoverheid.

OECD. 2009. “OECD Health Data - Version: November 09”. http://www.ecosante.org/index2.php?base=OCDE&langs=ENG&langh=ENG. Viewed 22 June 2010.

———.2010. “OECD Health Data - Version: June 10”. http://www.oecd.org/document/16/0,3746,en_2649_37407_2085200_1_1_1_37407,00.html. Viewed 3 July 2010.

Todosijevic, B., Aarts, K., and van der Kaap, H. 2010. Dutch Parliamentary Election Studies 1971-2006: Cumulative data-set and documentation. DANS Data Guide, DANS – *Data Archiving and Networked Services*, Volume 7, The Hague, p.264 (2010). http://povb-ecpr.org/node/89.

Sorenson, C., Drummond, M., and Kanavos, P. 2008. Ensuring value for Money in Health Care: the role of HTA in the European Union. Cornwall: World Health Organization 2008, on behalf of the European Observatory on Health Systems and Policies.

Stolk, E. A., and F. F. Rutten. 2005. The "health benefit basket" in the Netherlands. *Eur J Health Econ* Suppl:53-7.

# 7 Empirical analysis of HAS coverage decisions

Adalimumab is indicated for the treatment of ankylosing spondylitis[22], and was reviewed by NICE, SMC, CVZ and HAS to assess whether it should be funded by the healthcare system. HAS provided the following guidance:

> The Transparency Committee considers that HUMIRA provides the same improvement in actual benefit (ASMR II) as the other TNF antagonists (etanercept and infliximab) in the treatment of severe, active ankylosing spondylitis in adults responding inadequately to conventional treatment. (HAS 2006 section 4.2)

In contrast to NICE, SMC and the CVZ, HAS did not apply an explicit restriction to the use of adalimumab, nor a formal stopping rule, and in addition, awarded the compound with an ASMR of 'II'. The ASMR – *Amelioriation du Service Medicale Rendu*– reflects the incremental medical improvement the compound brings versus currently available therapies in France. Compounds are given a score between I and V; I being a compound that brings exceptional medical value, and V a compound that is judged to bring no incremental medical value. An ASMR of II reflected the HAS Committee's view that adalimumab has the potential to provide high medical value to patients with ankylosing spondylitis. Given the HAS coverage decision for adalimumab, what is the impact of evidence, process and context factors on HAS decisions? Is the difference in recommendation observed in this single case study a signal that the factors driving decision-making within the HAS appraisal process are different to those driving NICE, SMC or CVZ decisions? Or is this particular case study simply an outlier?

This chapter provides empirical analyses of coverage decisions made by HAS during the period 2004-2009. First, an overview of HAS, its objectives and appraisal process is provided. The methods for the analysis are then outlined, building upon the methods discussed in Chapter 3. Following this, the results of these multivariate analyses on HAS

22 Adalimumab (Humira) is an anti-inflammatory medicine and is indicated, among other diseases, for the treatment of adults with severe active ankylosing spondylitis (a disease causing inflammation and pain in the joints of the spine) who have not responded adequately to other treatments (European Medicines Agency 2009 EPAR Adalimumab)

coverage decisions are reported and explored, and limitations considered. The chapter concludes with a brief discussion about the empirical analyses performed for HAS.

## 7.1 HAS Appraisal Process

The Haute Autorité de Sante (HAS) in France was founded in 2004[23] as an independent scientific body, with the over-arching aim to support solidarity in the healthcare system and to reinforce the quality of treatment for the benefit of patients (HAS 2004). Among its many activities, HAS gives its advice on the therapeutic value of technologies reimbursed by the healthcare system. This advice is generated by a specific committee known as the Transparency Commission (*Commission de la Transparence*) whose main objective is to evaluate the therapeutic benefit of technologies newly licensed by a regulatory agency and for which there is a formal request for reimbursement that is submitted by the manufacturer to HAS. The Commission can appraise technologies for several purposes: i) to advise on whether to include the technology in the reimbursement list; ii) to give a new assessment due to a modification in the license of the technology (e.g. new indication); or iii) due to a renewal of reimbursement. In addition, upon request by the Minister of Health the Commission can evaluate whether or not a technology should be maintained on the reimbursement list. In order for a technology to be reimbursed, it must be included in the positive list known as the *Liste de Specialites Pharmaceutiques Remboursables aux Assures Sociaux* (Bellanger et al., 2005).

The Transparency Commission includes 31 members (Zentner et al. 2005) and is composed of physicians, pharmacists and specialists in epidemiology and research methods, who may serve on the Commission for a period of three years (HAS 2011). The appraisal also includes support from another group within HAS known as the *Direction de l'Evaluation Medicale, Economique et de Sante Publicque* (DEMESP). In addition, if needed, external experts may also be called to support the evaluation process. However, HAS does not operate a third party appraisal process like NICE and its MTA procedure.

There are several criteria upon which the appraisal is based (Secretariat General de la Commission de la Transparence 2005). Firstly, the evaluation is performed for each indication separately. For instance, even if a technology is indicated for two cancer types, the

[23] Decret n°2004-1398 du 23 décembre 2004.

appraisal process will treat each separately. The recommendation of the Commission is based on an assessment of the medical service rendered by the technology ('*Service Medical Rendu*'- SMR), by taking into consideration various factors including the severity of the disease, the efficacy and safety profile demonstrated, and the importance of the management of this disease from a public health standpoint. In addition, the Commission makes a recommendation on the incremental medical service rendered (*l'Amelioriation du Service Medical Rendu*, -ASMR) defined by examining the efficacy and safety profile of the technology relative to a specified comparator. There are five levels of ASMR as defined in the Commission's regulations, ranging from ASMR I which represents technologies that bring highly significant incremental medical value, and ASMR V which represents technologies that show no incremental medical value. The Commission also considers the target population for the technology and, for example, the nature of the technology in terms of its duration of treatment and dosing. It is important to note that the cost of the technology is not considered during the appraisal process and the price is not known during the appraisal by the Transparency Commission.

Once an appraisal is completed, the Commission's conclusion is transmitted to three parties: the Ministry of Health, the manufacturer and the *Comite Economique des Produits de Sante* (CEPS). This last committee is responsible for setting the price of medicines through negotiation with the manufacturers. These negotiations are based on several components, although they are primarily driven by the ASMR rating which represents the incremental benefit of the technology compared to the standard of care. Additional factors taken into consideration in price negotiations include the size of the target population, the manufacturer's research expenditure and advertising costs (Sorenson et al. 2008; Bellanger et al. 2005; Sandier et al. 2004). An ASMR of I-III allows the technology to obtain a premium price versus the comparator defined in the appraisal. An ASMR of IV defines a situation in which price parity price in relation to the comparator is implemented. An ASMR of V means that, per legislation, the technology cannot be included on the reimbursement list as it does not show any incremental benefit versus the comparators. The only condition upon which it could be included is if it provides cost-savings to the healthcare system. The National Union of Health Insurance Funds (*Union Nationale des Caisse d'Assurance Maladie – UNCAM*) has the authority to formally place a technology on a positive list, although the Ministry of Health and Social Security has the final say (Bellanger et al. 2005, Sorenson et al. 2008). Inclusion on the positive list is for a 5 year period after which there is a re-evaluation.

The clinical criteria considered by the Transparency Commission in its appraisal include information on the characteristics of the technology, particularly the indication and composition, information on the marketing authorisation granted in terms of dose, and contra-indications both within the EU and for marketing authorisations granted outside of the EU. To establish the therapeutic profile of the technology, an analysis of the clinical efficacy data is performed, vis á vis the formally identified comparator. Safety data are also reviewed and compared with the comparator. Priority is given to data emerging from randomised control trials, although the appraisal process also includes consideration of non-randomised observational data. In order to establish the medical service rendered by the technology, review of the seriousness of the disease is performed, and note is taken of the availability of alternatives and the need for treatment from a public health standpoint. The target population is established by considering the prevalence of the disease or condition in the French population and also includes, where appropriate, the estimation of the target population that would most particularly benefit from the treatment.

## 7.2 Methods

The overall objective was to examine the factors that influence decisions made by HAS to recommend, restrict or not recommend new technologies for use in the healthcare system. In addition to the general aims described in Chapter 1, the particular hypotheses relevant for the analysis of coverage decisions by HAS are highlighted in Box 7.1. Building on from the methods described in Chapter 3, this section describes the methods used to select the sample for analysis, the outcome variable and explanatory variables considered, and the statistical techniques adopted.

Box. 7.1 In light of the discussions outlined in Chapter 2, HAS-specific research objectives were to test whether:

- Clinical evidence variables (and hence 'proof' of degree of medical service rendered) impact on ASMR ratings – such that technologies achieving ASMR ratings I-II are accompanied by higher quality of evidence than those technologies achieving ASMRs of III or below.
- The political context, assessed by examining whether an election took place in the year of appraisal, impacts significantly on HAS decision-making by favouring higher ASMR ratings (I-II vs. III-IV or V)
- The pattern of coverage decision-making by HAS is changing over time– it is hypothesised that the proportion of ASMR V ratings is increasing over time relative to ASMR I-II ratings.
- The ASMR ratings reflected the nature of the disease for which the technology is indicated – therapeutic areas where unmet medical needs can be hypothesised to be high (e.g. rare diseases and/or cancer) positively on the odds of ASMRI-II versus ASMR III-IV or versus ASMR V.

### *7.2.1 Sample*

The drug technology appraisals performed by HAS formed the basis for the sample included in this analysis. The sample included all drug technology appraisals (as opposed to medical devices or other interventions) made during 2004-2009 indicated for an adult population (≥18 years). Technology appraisals were excluded from the analysis for any of the following reasons: i) they focused on a non-adult population; ii) they appraised non-drug interventions; iii) marketing authorisation was withdrawn; iv) the ASMR was not reported (HAS only); or v) the full guidance was not available.

Additional inclusion criteria were employed for this analysis of HAS which has numerous responsibilities, one of which is the provision of advice on new technologies available for patients. In total, the Transparency Commission issued more than 2000 recommendations in 2004-2009. Given the resource constraints available, it was not possible to review all 2000 recommendations to identify those of relevance for this research (i.e. not all recommendations provide ASMR, some recommendations are related to new mode of administration, new safety information or a re-review of technologies licensed prior to 2004 etc). In order to extract a relevant sample for this research, a list of technologies included in the SMC and NICE appraisals was created, and all the HAS recommendations linked to these technologies were extracted for review. The benefit of this approach was that it increased the opportunity for comparability across agencies by collecting information on a common list of compounds, and secondly it facilitated the streamlining of data extraction to those appraisals relevant for the research question.

### *7.2.2 Outcome variable*

To address the research question, HAS decisions were analysed through considering HAS outcomes in three categories, where the new technology can be classified as bringing of:

- Significant incremental medical benefit (ASMR I-II)
- Modest incremental medical benefit (ASMR III-IV)

or

- No incremental medical benefit (ASMR V)

The Transparency Commission's main role is to ascertain the incremental medical improvement that the technology brings versus currently available therapies in France through the ASMR. The level of funding, specifically its price and volume, is dependent upon the ASMR rating. For the purposes of this analysis, the five ASMR categories were 'collapsed' into three categories. Those decisions with an ASMR of I or II were combined in one category (equivalent to 'recommend' in the analyses of other HTA bodies), those with an ASMR III or IV were combined to create a second category (equivalent to 'recommend with restrictions'), and those decisions concluding with an ASMR of V were considered as the third category (equivalent to 'not recommended'). An alternative categorisation of the outcome variable was tested in a sensitivity analysis.

### *7.2.3 Explanatory variables*

In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, the HAS dataset includes 31 explanatory variables of interest. The first set of variables was related to the technology itself – the nature of the clinical evidence available, disease characteristics, whether cost-effectiveness evidence was put forward, and if so, the characteristics of that evidence. The second set of variables captured information relating to the process by which the recommendation was issued. Finally, the third set of variables in the data extraction form captured information on the context in which the guidance was issued (healthcare system, economic and social context), and thus aimed to capture information on variables linked to French health policy, and socio-economic status.

HAS differed from the three other HTA bodies in that no economic criteria were used by the Transparency Commission in its assessment of the technology. Therefore, no economic related variables were extracted for HAS. However, the clinical variables and variables related to the process and context of decision-making were extracted in a similar fashion to those extracted for NICE, SMC, and CVZ. In addition, for HAS, a set of specific variables were collected to reflect the specificities of the appraisal process. The first is related to the request for appraisal – that is, information was collected on the rationale for the request for review made by the manufacturer, including whether it was a request to review a new technology, or a new indication of an existing technology, or because it was a class review, or a re-review etc. This is a specific aspect of the HAS reimbursement review system for which relevant information was collected. In

addition, information was extracted about whether the technology was intended for inpatient or outpatient use, similar to what was collected for the CVZ. The rate of reimbursement was collected, which is determined by specific legislation and is driven primarily by the chronic nature and severity of the disease and on the profile of the technology, but not by the ASMR. With regard to post-reimbursement commitments, data was extracted on whether a request was made for an observational study to be conducted, featuring the newly reimbursed technology.

### *7.2.4 Data extraction form*

The data set was created by extracting data from appraisal reports/documents made available on the HAS website . The process of data extraction followed the protocol set out in Chapter 3. Table 7.1 provides the list of variables extracted to create the HAS dataset, as well as the accompanying decision rules and definitions.

**Table 7.1 HAS dataset: Included Variables, their Definition, Data Extraction Rule and Data Sources**

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **1** | Number of RCTs considered in decision | Count | The number of distinct Randomised Controlled Trials (RCTs) that provide data related to the therapeutic indication under evaluation.<br>Excluded: studies that are single arm, that have no randomisation, or that are non-interventional. | Advice from the Commission de la Transparence, section 3 (3.1 – 3.3) |
| **2** | Size of population included in RCTs | Numeric | Mean number of patients per RCT. | Advice from the Commission de la Transparence, section 3 (3.1 – 3.3) |
| **3** | Length/extent of follow-up in RCT | Numeric | Mean number of weeks that data is collected on patients that entered the RCTs (see variable no. 1). | Advice from the Commission de la Transparence, section 3 (3.1 – 3.3) |
| **4** | Statistically Significant results | Categorical (yes/no/inconsistent) | Presence of statistically significant superiority of technology versus the comparator for primary endpoint(s).<br><br>If more than one RCT was considered, and the technology showed statistically significant superiority in one trial, but not in another, the results were considered to be 'inconsistent' and classified as such. RCTs designed as 'non-inferiority' studies were classified as not showing any statistically significant superiority (i.e. 'no'). | Advice from the Commission de la Transparence, section 3 (3.1 – 3.3) |
| **5** | Relevance of RCT to payer decision | Numeric | Percentage of RCTs where active comparator was used. | Advice from the Commission de la Transparence, section 3 (3.1 – 3.3) |
| **6** | Number of observational studies considered in guidance | Count | Number of observational studies providing information to support study drug. Observational studies in this circumstance are defined as studies that are non interventional (i.e. do not explicitly request the patient to take particular medication or the physician to follow particular protocol). | Advice from the Commission de la Transparence, section 3 (3.1 – 3.3) |
| **7** | Priority disease area | Categorical – yes/no | This variable aims to capture the health policy context in which the payer decision is made, by capturing whether the pharmaceutical in question is linked to a disease area that is prioritized by the ministry of health. Priority disease areas were identified by examining government plans/health documents that highlight national health care system focus. | CNRS (2004) |
| **8** | Orphan Status | Categorical – yes/no | This variable captured information on whether or not the technology was recognized by the European Medicines Agency (EMA) as an orphan designated medicine. | European Medicines Agency (accessed 2010-2011); Cover Page of Avis from the Commission de la Transparence |
| **9** | Therapeutic Area | Categorical – 13 | The British National Formulary (BNF) categories were used to classify | British National Formulary (2010) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| | | categories | each technology into the corresponding therapeutic area. | |
| **10** | Prevalence of disease/clinical condition | Numeric | Reported number of patients eligible for treatment, as per the Summary Product Characteristics and indication of the medication under evaluation. | Advice from the Commission de la Transparence, section 4.4 |
| **11** | Availability of alternative therapies in current treatment setting. | Categorical – yes/no | An alternative was considered to be available if comparators were clearly defined in the review by the HTA agency. An alternative was considered NOT to be available if it was stated as such in the appraisal, or if 'best supportive care' or 'palliative care' was specified as the comparator. | Advice from the Commission de la Transparence, sections 2.2 and 2.3, and 4.3 |
| **12** | Consideration of Cost Utility Analysis in guidance | Categorical – CUA performed or no CUA | Presence or absence of a cost-utility analysis. | Secretariat General de la Commission de la Transparence (2005) |
| **13** | Inclusion of patient submission | Categorical – yes/no | A patient submission was considered to have been included as part of the appraisal process if a submission from a patient group was acknowledged within the appraisal document (e.g. SMC, CVZ) or posted on the webpage pertaining to the guidance (NICE). | Secretariat General de la Commission de la Transparence (2005) |
| **14** | Number of Decision Makers Accountable | Numeric | Captures the number of decision-makers accountable for guidance issued, as reported. For NICE, this information was extracted from Appendix B of each guidance, for the SMC, from the minutes of the meeting. For the CVZ and HAS, directly from the description of the committee webpages. | Zentner et al., 2005; Secretariat General de la Commission de la Transparence (2005) |
| **15** | Cost-effectiveness evaluation component in process | Categorical – yes/no | Captures whether cost-effectiveness is a component of the decision-making process or not. If cost-effectiveness analysis is a formal part of the appraisal process, this variable was marked as 'yes'. | Secretariat General de la Commission de la Transparence (2005) ; Sorenson et al. (2008) |
| **16** | Budget impact as a component of decision-making process | Categorical – yes/no | Captures whether budget impact considerations are part of decision-making process | Secretariat General de la Commission de la Transparence (2005) ; Sorenson et al. (2008) |
| **17** | Pricing known during appraisal process | Categorical – yes/no | Captures whether the price of the technology under appraisal was known during the assessment. | Secretariat General de la Commission de la Transparence (2005) |
| **18** | Number of technologies appraised simultaneously | Count | This variable captures the number of technologies appraised simultaneously in the appraisal. | Secretariat General de la Commission de la Transparence (2005) |
| **19** | Different process for medications destined for hospital or retail use | Categorical – yes/no | Records whether funding decisions for medications follow different processes depending on whether they are destined for hospital or retail prescription. | Secretariat General de la Commission de la Transparence (2005) |
| **20** | Accountability of drug budget | Categorical – yes/no | The HTA agency was examined to assess whether or not the agency making the funding decisions is also accountable for the drug budget. | Sorenson et al. (2008) |
| **21** | Independence of decision-making agency | Categorical – yes/no | This pertains to whether the HTA body is independent of Ministry of Health or part of it. | Sorenson et al. (2008) |

| # | Variable Descriptor | Unit measure | Definition | Data Sources |
|---|---|---|---|---|
| **22** | Date guidance was issued | Numeric | Year when coverage decision was issued | Cover Page of Avis from the Commission de la Transparence |
| **23** | Population size – Agency coverage | Numeric | Estimate of population size within remit of the agency performing the evaluation. | Eurostat (2010) |
| **24** | GDP-healthcare expenditure | Numeric (%) | Percentage of GDP spent on healthcare, during year of decision | OECD (2009) |
| **25** | Healthcare expenditure on pharmaceuticals | Numeric (%) | Percentage of healthcare budget spent on pharmaceuticals per patient per year, during the same year in which the appraisal was published. | Econ-Sante` France (2010) ; IRDES (2009) |
| **26** | Drug funding process within healthcare system – whether centralized or decentralized | Categorical – centralized, decentralized | States whether drug funding process within the healthcare system is centralized at a national level or whether funding decisions are decentralized to the regional level | Sorenson et al. (2008) |
| **27** | Election year at time of decision | Categorical – yes/no | This variable captures whether payor decision was made within an election year. An election year was defined as a year in which either national government or regional elections took place. | Le ministre de l'intérieur, de l'outre-mer, des collectivités territoriales et de l'immigration (2004) and (2007) |
| **28** | Inpatient Use | Categorical – yes/ no | This variable provides information on whether or not the technology was specified for use within an inpatient setting. | Cover Page of Avis from the Commission de la Transparence and section 4.5 |
| **29** | Post-approval Study request | Categorical – yes/ no | This variable provides information on whether reimbursement was granted with the condition that real-life observational data on the technology would be provided within a specified time period. | Advice from the Commission de la Transparence, section 4.4 |
| **30** | Reason for Reimbursement request | Categorical – 5 categories | Collected for HAS only – captures the rationale for the reimbursement request – including whether the review was for a new technology, a technology with an indication extension, a class review, a modification of mode of administration/dosage. Extracted directly from first page of appraisal report. | Cover Page of Avis from the Commission de la Transparence |
| **31** | Reimbursement Level | Numeric (%) | For HAS only. This captures information on the level of reimbursement granted for the technology as per national legislation which attributes levels of reimbursement based on characteristics of the compound. This information was extracted from the last section of the HAS appraisal report. | Advice from the Commission de la Transparence, section 4.5 |

### *7.2.5 Statistics*

The methods for the descriptive statistics and multivariate analyses were described in Chapter 3. Descriptive statistics were calculated for each extracted variable, stratified by outcome group (recommended, restricted or not recommended). Following a descriptive analysis of the dataset, a multinomial logit regression was modelled. In the multivariate analysis, the base outcome utilised was the 'AMSR I-II' outcome. The objective of this analysis was to obtain a parsimonious model that best reflected the main drivers of HAS decision-making.

## 7.3 Results

### *7.3.1 Sample characteristics*

A total of 351 technology appraisals performed between January 2004 and June 2009 were retrieved from the HAS website. Of these, 315 full submissions were included for analysis. 36 drug reviews were excluded from the analysis for the following reasons: either the ASMR was not reported (n=20) or the review focused on a non-adult population (n=16).

Table 7.2a shows the HAS guidance issued between 2004- June 2009, and the distribution of drugs by ASMR rating. Within this five-year period, of the 315 HAS decisions analysed, 3% of decisions awarded an ASMR of 1, meaning that the technology was considered to bring highly significant medical benefit, and 15% of decisions awarded an ASMR of II, in instances where the committee considered that the technology would bring significant medical improvement. The majority of decisions (44%) concluded that there was no medical improvement associated with the technology (ASMR V). Table 7.2b shows the proportion of technologies in ASMR I-II, ASMR III-IV and ASMR V categories.

**Table 7.2a Outcome of HAS guidance issued between 2004-June 2009 – HAS decisions stratified by ASMR**

| HAS guidance ASMR | Number of coverage decisions | Percentage |
|---|---|---|
| I | 11 | 3% |
| II | 45 | 15% |
| III | 60 | 19% |
| IV | 60 | 19% |
| V | 139 | 44% |
| **Total** | **315** | **100%** |

**Table 7.2b Outcome of HAS Guidance issued between 2004-June 2009 – HAS decisions stratified by ASMR category**

| HAS guidance | Number of coverage decisions | Percentage |
|---|---|---|
| ASMR I-II | 56 | 18% |
| ASMR III-IV | 120 | 38% |
| ASMR V | 139 | 44% |
| **Total** | **315** | **100%** |

### *7.3.2 Descriptive statistics*

Clinical Evidence

Descriptive statistics for HAS are summarized in Table 7.3. On average HAS took into consideration 2 RCTs per appraisal in its review process, with an average sample size of 1154 patients, and duration of 49 weeks. Decision outcomes differed in terms of the number of RCTs considered by the committee, as well as the duration. Perhaps unexpectedly, ASMRI-II interventions had a lower number of RCTs considered by the HTA committee (1.8 trials), compared to interventions that had an ASMR III-IV or ASMR V (2.4 and 2.5 respectively), and this difference between outcome categories was statistically significant (p=0.013). In addition, differences between outcome categories were observed in the duration of clinical trials considered: this ranged from 63 to 51 to 43 weeks for the ASMRI-II, ASMRIII-IV and ASMR V groups respectively (p=0.0013). While differences in the number and duration of clinical trials was observed, differences in the size of the clinical trials considered in the Transparency Commission's reviews was not statistically significant between outcome groups (ranging from 917 – 1199 patients).

Aside from the size and number and duration of RCTs, the results of the RCTs were also captured. The majority (73%) of technologies rated ASMRI-II for reimbursement by HAS were supported by RCTs demonstrating superiority of the technology with regard to the primary endpoint(s). This is in contrast to the ASMRIII-IV and ASMR V technologies, where the proportion of RCTs demonstrating superiority was lower (56% and 31% respectively). This difference between groups in terms of demonstration of superiority was statistically significant ($p<0.001$). The comparator used within the clinical trial programme was assessed – in particular, the percentage of comparisons made to 'active' comparators as opposed to placebo was recorded. Unexpectedly, interventions rated ASMRI-II had a lower number of comparisons to active agents (25%) than interventions that were ASMRIII-IV or ASMR V (40% and 50% of RCTs

with active comparators, respectively). The differences observed between the groups were statistically significant (p=0.0137).

Consideration of non-randomised observational data was also recorded. Overall, in the majority of drug reviews, observational data was not considered and there were no significant differences between outcome groups.

Economic Evidence

None of the HAS reviews included the evaluation of the economic characteristics of the technology. This is entirely driven by the HAS review process which does not incorporate such information in its reimbursement recommendation. Therefore, no economic variables are described.

Disease characteristics

The HAS review process requires information on the number of patients eligible for treatment with the drug under review. This ranged from 61,776 patients for ASMR I-II interventions to 246,940 and 907,741 patients in the ASMRIII-IV and ASMRV interventions. All statistical tests performed suggested that these differences were highly statistically significant.

The availability of alternative therapies was assessed to ascertain if it differed between ASMR I-II and ASMRIII-IV or ASMR V interventions. In the majority of technologies appraised, an alternative was available (in 89% of cases, all decisions considered). There were differences across the outcome groups which were statistically significant (p= 0.014).

**Table 7.3 HAS Coverage Decisions: Descriptive statistics for extracted variables, by coverage decision (ASMR I-II, ASMR III-IV, ASMR V)**

| Variable | Total (n=315) | | | ASMR 1 + 2 (n=56) | | | ASMR 3 + 4 (n=120) | | | ASMR V (n=139) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | |
| Number of RCTs considered in decision | 2.3 | 2.1 | 2.6 | 1.8 | 1.4 | 2.2 | 2.4 | 2.0 | 2.8 | 2.5 | 2.1 | 3.0 |
| Size of population included in RCTs | 1154 | 824 | 1484 | 917 | 293 | 1541 | 1194 | 632 | 1756 | 1199 | 690 | 1708 |
| Length/extent of follow-up in RCT (weeks) | 49 | 41 | 57 | 63 | 47 | 78 | 51 | 38 | 65 | 43 | 30 | 55 |
| Statistically Significant results - *yes (1)* | 47% | 33% | 43% | 73% | 38% | 65% | 56% | 37% | 55% | 31% | 19% | 33% |
| *no (0)* | 20% | 21% | 31% | 10% | 4% | 21% | 12% | 19% | 35% | 30% | 24% | 39% |
| *inconsistent (2)* | 33% | 12% | 20% | 18% | 0% | 14% | 32% | 5% | 15% | 38% | 18% | 32% |
| Relevance of RCT to payer decision | 42% | 37% | 48% | 25% | 13% | 38% | 40% | 31% | 49% | 50% | 42% | 59% |
| Number of observational studies considered in guidance | 0.3 | 0.2 | 0.4 | 0.5 | 0.0 | 1.0 | 0.3 | 0.1 | 0.6 | 0.2 | 0.1 | 0.3 |
| Consideration of Cost Utility Analysis in guidance | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Prevalence of disease/clinical condition | 511,047 | 314,122 | 707,972 | 61,776 | -31,135 | 154,687 | 246,940 | 118,391 | 375,489 | 907,741 | 489,394 | 1,326,088 |
| Social Perspective adopted | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Availability of alternative therapies in current | 89% | 85% | 92% | 81% | 70% | 92% | 86% | 80% | 92% | 94% | 90% | 98% |

| Variable | Total (n=315) | | | ASMR 1 + 2 (n=56) | | | ASMR 3 + 4 (n=120) | | | ASMR V (n=139) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | |
| treatment setting. | | | | | | | | | | | | |
| Inclusion of patient submission | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Cost-effectiveness evaluation component in process | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Budget impact as a component of decision-making process | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Pricing and Reimbursement decided jointly | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Date guidance was issued | 2007 | 2006 | 2007 | 2006 | 2006 | 2006 | 2006 | 2006 | 2007 | 2007 | 2007 | 2007 |
| Population size – Agency coverage (thousands) | 63,400 | 63,300 | 63,500 | 63,200 | 64,400 | 63,000 | 63,300 | 63,200 | 63,500 | 63,500 | 63,400 | 63,600 |
| GDP-healthcare expenditure | 11% | 11% | 11% | 11% | 11% | 11% | 11% | 11% | 11% | 11% | 11% | 11% |
| Healthcare expenditure on pharmaceuticals | 519 € | 517 € | 522 € | 512 € | 506 € | 518 € | 517 € | 512€ | 521 € | 524 € | 520 € | 527 € |
| Election year at time of decision | 30% | 25% | 36% | 30% | 18% | 43% | 38% | 30% | 47% | 24% | 17% | 31% |
| Priority disease area | 70% | 64% | 75% | 71% | 59% | 84% | 73% | 64% | 81% | 66% | 58% | 74% |
| Orphan Designated | 9% | 6% | 12% | 23% | 12% | 35% | 12% | 6% | 17% | 1% | -1% | 3% |

| Variable | Total (n=315) | | | ASMR 1 + 2 (n=56) | | | ASMR 3 + 4 (n=120) | | | ASMR V (n=139) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | |
| Reimbursement Level | 70% | 68% | 73% | 73% | 68% | 79% | 74% | 70% | 77% | 67% | 64% | 70% |
| Hospital use only | 51% | 46% | 57% | 74% | 61% | 86% | 56% | 46% | 65% | 39% | 31% | 48% |
| Request for post-marketing study | 24% | 19% | 28% | 33% | 20% | 47% | 30% | 21% | 38% | 15% | 9% | 21% |
| Reason for HAS review | | | | | | | | | | | | |
| *Inscription on social security/reimbursement list* | 44% | 38% | 49% | 36% | 23% | 49% | 48% | 38% | 57% | 44% | 36% | 52% |
| *Re-evaluation of previous appraisal* | 8% | 10% | 23% | 7% | 0% | 28% | 8% | 5% | 25% | 9% | 9% | 29% |
| *Extension of indication* | 34% | 86% | 118% | 36% | 68% | 146% | 28% | 60% | 110% | 38% | 90% | 139% |
| *Renewal of inscription on social security* | 13% | 37% | 67% | 21% | 41% | 130% | 14% | 31% | 82% | 9% | 16% | 53% |
| *New reimbursement criteria* | 1% | -1% | 10% | 0% | 0% | 0% | 3% | -2% | 27% | 0% | 0% | 0% |
| BNF Category | | | | | | | | | | | | |
| *cardiovascular system* | 14% | 10% | 18% | 18% | 8% | 28% | 13% | 7% | 20% | 13% | 7% | 19% |
| *central nervous system* | 18% | 14% | 23% | 2% | -2% | 5% | 20% | 13% | 27% | 24% | 17% | 31% |
| *ear, nose and oropharynx* | 0% | 0% | 1% | 0% | 0% | 0% | 0% | 0% | 0% | 1% | -1% | 2% |
| *endocrine system* | 6% | 4% | 9% | 4% | -1% | 9% | 7% | 2% | 11% | 7% | 3% | 12% |
| *eye* | 2% | 0% | 3% | 2% | -2% | 5% | 2% | -1% | 4% | 2% | 0% | 5% |
| *gastro-intestinal system* | 3% | 1% | 6% | 2% | -2% | 5% | 2% | -1% | 4% | 6% | 2% | 10% |
| *infections* | 12% | 8% | 15% | 2% | -2% | 5% | 13% | 7% | 20% | 14% | 8% | 20% |

| Variable | Total (n=315) | | | ASMR 1 + 2 (n=56) | | | ASMR 3 + 4 (n=120) | | | ASMR V (n=139) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | |
| *malignant disease and immunosuppression* | 24% | 19% | 29% | 39% | 26% | 52% | 26% | 18% | 34% | 17% | 10% | 23% |
| *musculoskeletal and joint diseases* | 9% | 6% | 12% | 25% | 13% | 37% | 8% | 3% | 12% | 4% | 1% | 8% |
| *nutrition and blood* | 3% | 1% | 6% | 7% | 0% | 14% | 3% | 0% | 5% | 3% | 0% | 6% |
| *obstetrics, gynaecology, and urinary-tract disorders* | 0% | 0% | 1% | 0% | 0% | 0% | 0% | 0% | 0% | 1% | -1% | 2% |
| *respiratory system* | 2% | 1% | 4% | 0% | 0% | 0% | 2% | -1% | 4% | 4% | 0% | 7% |
| *skin* | 4% | 2% | 7% | 0% | 0% | 0% | 6% | 2% | 10% | 5% | 1% | 9% |

For each drug review, in order to assess whether the type of disease impacted on decision making, the disease was coded using BNF categories – 13 categories are used ranging from cardiovascular and central nervous system to infections, and skin diseases. Across all decisions, the majority of decisions were linked to medications for malignant disease and immunosuppression, central nervous system, and cardiovascular disease and infections. This differed, however, between the decision categories: for example, technologies for the treatment of malignant disease and immunosuppression appeared to be more likely to fall in the ASMR I-II category (39%) than in the ASMR III-IV (26%) or ASMR V groups (17%). The most common disease category in the ASMR I-II group was therapies linked to malignant disease and immunosuppression (39% of recommendations made for this disease category). It was also the most common disease area in the ASMR III-IV group. In the ASMR V group, technologies were most frequently linked to the management of diseases related to the central nervous system (24%). The differences observed between outcomes groups in terms of the distribution of decisions by BNF categories were statistically significant ($p \leq 0.001$).

Orphan designation is given to specific therapies by the EMA if they fulfil specific criteria, namely related to the prevalence of the disease, the clinical profile of the disease and level of unmet need. During 2004-2009, 9% of reviewed drugs had an orphan designation. A higher proportion of orphan medicines were in the ASMR I-II group (23%) compared to the 'ASMR III-IV' (12%) and 'ASMRV' interventions (1%). These differences between outcome groups were statistically significant ($p \leq 0.001$).

<u>HAS assessment process characteristics</u>

A series of variables related to the decision-making process were recorded. Firstly, the HAS review process does not foresee input from patient groups. Secondly, with regard to economic considerations, the HAS review process does not incorporate economic considerations in its decision making – therefore, it was not possible to include neither cost-effectiveness/cost-utility analyses, nor budget impact estimates in this analysis. Finally, it is not part of the HAS appraisal process to decide on funding and pricing decisions jointly – indeed, once the Transparency Commission has given its decision, this information is provided to the Economic Committee (*Comittee Economique)* which negotiates the price and volume with the respective manufacturer. In general, the HAS

reviews single technologies, although there are a few cases where class reviews are performed, in which case several technologies may be reviewed simultaneously.

The HAS not only reviews new technologies for reimbursement purposes, but also reviews decisions either after a specific time period, or if there is a new indication, or a class review. Therefore, there are a number of rationales for HAS to conduct a review. The most frequent reason, occurring in 44% of cases, is the so called 'Inscription Securite Sociale et Collectivites', related to the registration of the technology on the reimbursed list, or positive list. The second most common reason (34%) is related to a review due to an expansion of the indication/use of the technology. The reasons for HAS appraisal were not statistically significantly different between the ASMR groups.

Related to the HAS review process is whether the technology is considered to be available for use exclusively in a hospital setting or outpatient setting. Of the reviewed technologies, approximately 52% of ASMRIII-IV were for hospital use only. This differed between the outcome groups – more of the ASMRI-II interventions were for hospital use only (76%) compared to the restricted – ASMR III-IV - (56%) or not reimbursed – ASMR V- interventions (39%). These differences are statistically significant (p=0.0001).

As part of its process the HAS also specifies the level of reimbursement for the technology – this is governed by Ministry of Health criteria based on the disease area, whether the disease is chronic, if it is predominantly a paediatric condition or a condition prevalent in the elderly etc. Based on this algorithm, the technology is associated with a reimbursement level. Such information was collected as part of this research. The average reimbursement rate is 70%, across the 315 appraisals; ranging from approximately 73% in the ASMRI-II and ASMRIII-IV groups to 66% in the ASMR V group[1]. The difference observed in the mean level of reimbursement between the ASMRIII-IV and ASMR V interventions is statistically significant (p=0.0082)

One of the characteristics of the HAS process is the possibility for the Transparency Commission to request an observational study to ascertain the effectiveness/use/safety of the technology once used within the French healthcare system. Of the HAS reviews

[1] It should be noted that ASMR V medications cannot be inserted in the reimbursement list unless they are associated with lower treatment costs than standard care.

considered, in 24% of cases HAS requested these post-launch observational studies. There were more requests made for such research for ASMRI-II and ASMRIII-IV interventions than for interventions considered to have no incremental medical benefit (ASMR V). These differences were statistically significantly different (p=0.004).

Finally, as part of the HAS appraisal process, the Commission can indicate whether the reimbursement should be granted for the indicated population, a subset of the population or for none of the population. The last scenario is very rare (1% of reviewed decisions). The most common is that the HAS recommends use for the entirety of the indicated population (71%) of cases. However, this varies between the outcome groups. Those interventions that have an ASMR I-II, the use of the technology for the entirety of the population is granted in 91% of cases, as opposed to 66% and 69% of cases for interventions that are rated ASMRIII-IV or ASMR V, respectively. The differences observed were statistically significant (p= 0.009).

Socio-economic context of HAS decision-making

A series of variables were recorded to capture the socio-economic context in which HAS decisions were made. In particular, information was recorded on the year of the appraisal (mean 2007), size of population under HAS remit (approx. 63.4 million), percentage of GDP spent on healthcare (11%), and whether appraisal coincided with an election year (30% of cases). There was strong correlation between the year of appraisal and the other socio-economic factors. The year of appraisal, population size, healthcare expenditure on pharmaceuticals and percentage of decisions made within an election year, appeared to vary significantly between outcome groups. For example, during the election year, there were more recommendations (30%) and restrictions (38%) made than there were 'non recommendations' (24%). These differences were statistically significant (p=0.039).

Summary of descriptive analysis

In order to assess the factors that impact on funding decisions in France, 315 decisions made by the HAS in the period 2004-2009(June) were reviewed. These 315 appraisals correspond to the compounds that had been included in SMC and NICE reviews in the same period. The HAS recommendations were analysed according to the ASMR that was attributed to the technology for the specific patient population under review. The ASMR levels were grouped into three categories – ASMR I-II, ASMR III-IV, and

ASMR V. To identify those factors that can explain HAS decision-making patterns, information on 31 explanatory variables was collected for decisions made by HAS in 2004-2009.

Of these variables, descriptive analysis suggests that approximately half of the tested variables appear to be a contributing factor in determining HAS decision-making (Table 7.4). For these variables, statistically significant differences were observed between interventions that had an ASMR I-II, III-IV and V ($p \leq 0.05$). Highly statistically significant differences ($p \leq 0.01$) were observed for variables highlighted in bold.

**Table 7.4 HAS descriptive statistics: statistically significant variables ($p \leq 0.05$)**

| **Component 1** | **Variables** |
|---|---|
| Clinical Package | **No of RCTs**<br>RCT demonstrates superiority<br>Duration of RCT<br>Active comparator used |
| Economic Package | -- |
| Disease characteristics | **Prevalence of disease**<br>Availability of alternatives<br>**Orphan status**<br>Therapeutic area: cancer therapies, musculoskeletal/joint diseases, cardiovascular disease, central nervous system diseases, infectious diseases |
| **Component 2 – Process Characteristics** | Level of reimbursement<br>Hospital use only<br>Request for post-launch study<br>Reason for request |
| **Socio-economic Context** | Date of Review<br>Pharmaceutical Exp per patient<br>National Population size<br>**Election year** |

Note: variables in bold text were statistically significant at the $p \leq 0.01$ level

### *7.3.3 Multivariate analysis*

Following the model specification process described in Chapter 3 which included the development of a preliminary model (Appendix D), the base case model of HAS decision-making was developed which contains eight variables impacting significantly on HAS decisions, including clinical, disease and process related variables. The model yielded a pseudo R-squared value of 0.19, suggesting that it appears to explain approximately 19% of the variability observed in coverage decisions made by the HAS (Table 7.5). There were variables which had a statistically significant impact on both coverage options (ASMR I-II vs. ASMR III-IV or ASMR V): prevalence of disease; orphan designation; indicated for treatment of central nervous system (CNS) disorders,

infectious disease or musculoskeletal and joint diseases; and the absence of data on prevalence all impacted significantly on HAS coverage decision-making. Specifically, a unit increase in the prevalence of the target population, or an indication for the treatment of CNS disorders or infectious diseases increased the probability of an ASMR III-IV or ASMR V, while orphan designation, license for the treatment of musculoskeletal and joint diseases, and absence of information on the prevalence of the target population increased the log odds of an ASMR I-II. If the technology was indicated for the treatment of musculoskeletal and joint diseases, this decreased the log odds of an AMSR I-II. Oddly, the absence of prevalence data increased the odds of an ASMR I-II.

**Table 7.5 Multivariate analysis of HAS coverage decisions 2004-2009: base case model results (n=315)**

| ASMR III-IV | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -0.42 | 0.277 | -1.179 | 0.338 |
| Disease prevalence | 0.00000049 | 0.113 | -0.00000012 | 0.0000011 |
| Orphan designation status | -1.27 | 0.010 | -2.24 | -0.30 |
| Central nervous system | 2.87 | 0.007 | 0.79 | 4.95 |
| Infections | 2.31 | 0.036 | 0.15 | 4.47 |
| Musculoskeletal and joint diseases | -1.58 | 0.002 | -2.61 | -0.56 |
| Healthcare expenditure on pharmaceuticals | 0.0091 | 0.273 | -0.0072 | 0.025 |
| Missing data on prevalence of disease | -2.15 | 0.001 | -3.38 | -0.91 |
| Constant | -3.45 | 0.419 | -11.81 | 4.91 |
| ASMR V | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -1.35 | 0.001 | -2.17 | -0.53 |
| Disease prevalence | 0.00000088 | 0.005 | 0.00000027 | 0.0000015 |
| Orphan designation status | -3.52 | <0.001 | -5.18 | -1.86 |
| Central nervous system | 3.17 | 0.003 | 1.07 | 5.27 |
| Infections | 2.21 | 0.046 | 0.04 | 4.38 |
| Musculoskeletal and joint diseases | -2.58 | <0.001 | -3.87 | -1.28 |
| Healthcare expenditure on pharmaceuticals | 0.024 | 0.008 | 0.0063 | 0.042 |
| Missing data on prevalence of disease | -2.75 | <0.001 | -3.98 | -1.51 |
| Constant | -10.67 | 0.021 | -19.73 | -1.62 |

Note: Technologies with ASMR I-II are the reference case. Multinomial logistic regression, pseudo R-squared: 0.19.

There were other variables in the HAS model whose impact was observed in one of the two coverage alternatives. In comparison to an ASMR I-II, the demonstration of clinically significant superiority decreased the log odds of an ASMR V, but did not impact statistically significantly on the log odds of an ASMR III-IV. The demonstration of clinical superiority of the technology in a clinical trial statistically significantly decreased the log odds of an ASMR V relative to an ASMR I-II (p=0.001). A unit increase in pharmaceutical expenditure increased the odds of an ASMR V relative to an ASMR I-II (p=0.012), but did not demonstrate a statistically significant effect on the odds of ASMR III-IV.

Impact of alternative model specifications- sensitivity analyses

Sensitivity analysis was conducted on the HAS regression model. This included: i) examining the impact of a binary rather than three-category outcome variable; ii) restricting the base case analysis to observations without missing entries, thus excluding observations with imputed values; iii) an alternative categorization of the ASMR is used, and (iv) examining the impact of assuming ordinality of the coverage decision.

In the first sensitivity analysis the ASMR ratings were collapsed into two categories: those technologies showing at least some incremental therapeutic benefit (ASMR I-IV) and those showing no incremental therapeutic benefit (ASMR V). This binary outcome variable was regressed with each explanatory variable individually to ascertain whether those variables that impacted statistically significantly in the base case analyses were still significantly impacting on the binary coverage decision. This sensitivity analysis using a binary outcome variable includes seven variables, and obtains a R-squared of 0.18, suggesting that it can explain approximately18% of variation in HAS decision-making, similar to the base case model with a pseudo R-squared of 0.19. A set of variables remained common in both models: demonstration of superiority, the prevalence of the condition and lack of prevalence data, orphan designation, treatment for central nervous system diseases and musculoskeletal/joint diseases. In this sensitivity analysis, technologies indicated for infectious diseases were no longer significant predictors of HAS decision-making.

In this sensitivity analysis, three variables increased the odds of an ASMR V, namely the prevalence of the disease, the nature of the disease, and pharmaceutical expenditure. A unit increase in the prevalence of the licensed indication increased the odds of an ASMR V ($p=0.006$), as did a license for the treatment of central nervous system conditions ($p=0.058$). A unit increase in the mean national drug expenditure per patient also increased the odds of an ASMR V ($p=0.003$).

Conversely, a combination of clinical and disease variables increased the odds of an ASMR I – IV. Demonstration of superiority in efficacy within the RCT, increased the odds of an ASMR $\geq$ IV ($p=0.019$). Technologies licensed for the treatment of musculoskeletal/joint conditions increased the odds of an ASMR $\geq$ IV, and orphan designation also significantly increased the odds of an ASMR $\geq$ IV. This was consistent

with the effect also observed in the base case analysis. Oddly, and consistently with what was observed in the base case analysis, the lack of prevalence data increased the odds of an ASMR ≥ IV.

**Table 7.6 Multivariate analysis of HAS coverage decisions 2004-2009: sensivitiy analysis using binary outcome variable (ASMR ≥ IV vs. ASMR V) (n=315)**

| Odds of ASMR ≥ IV | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | 1.03 | <0.001 | 0.48 | 1.57 |
| Disease prevalence | -0.00000046 | 0.006 | -0.00000078 | -0.00000013 |
| Orphan designation status | 2.63 | 0.001 | 1.13 | 4.12 |
| Central nervous system | -0.62 | 0.058 | -1.27 | 0.022 |
| Musculoskeletal and joint diseases | 1.51 | 0.006 | 0.42 | 2.59 |
| Healthcare expenditure on pharmaceuticals | -0.017 | 0.003 | -0.029 | -0.0058 |
| Missing data on prevalence of disease | 1.08 | 0.007 | 0.30 | 1.86 |
| Constant | 8.79 | 0.004 | 2.78 | 14.80 |

**Table 7.7 Multivariate analysis of HAS coverage decisions 2004-2009: sensivitiy analysis excluding incomplete observations from the sample (n=235)**

| ASMR III/IV | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -1.07 | 0.049 | -2.13 | -0.0069 |
| RCT duration of follow-up | -0.0051 | 0.123 | -0.012 | 0.0014 |
| Disease prevalence | 0.000010 | 0.093 | -0.000002 | 0.000023 |
| Orphan designation status | -1.65 | 0.009 | -2.89 | -0.42 |
| Musculoskeletal and joint diseases | -2.30 | 0.001 | -3.63 | -0.97 |
| Healthcare expenditure on pharmaceuticals | 0.014 | 0.199 | -0.0074 | 0.035 |
| Constant | -4.95 | 0.375 | -15.88 | 5.99 |
| ASMR V | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -2.17 | <0.001 | -3.29 | -1.04 |
| RCT duration of follow-up | -0.011 | 0.012 | -0.02 | -0.0023 |
| Disease prevalence | 0.000011 | 0.078 | -0.0000012 | 0.000023 |
| Orphan designation status | -3.44 | <0.001 | -5.29 | -1.59 |
| Musculoskeletal and joint diseases | -2.73 | 0.001 | -4.38 | -1.08 |
| Healthcare expenditure on pharmaceuticals | 0.035 | 0.003 | 0.012 | 0.058 |
| Constant | -15.20 | 0.012 | -27.05 | -3.36 |

Note: Technologies with ASMR I or II are the reference case. Multinomial logistic regression, pseudo R-squared: 0.23.

The second sensitivity analysis restricted the analysis to the sample of observations without missing entries. Thus, imputation was not used in this analysis, and the total sample covered 235 observations (n=315 in base case analysis). The regression model on the subset of complete observations results in a pseudo R-squared of 0.23, suggesting it can explain approximately 23% of variability in HAS decision making. This figure is higher than observed in the base-case model, where the pseudo R-squared was 0.19. The sensitivity analysis regression model confirmed the important impact of several variables including: the demonstration of superiority in the clinical trial which increased the probability of an ASMR I-II, the prevalence of the target population, orphan designation, and treatment of musculoskeletal/joint diseases. The impact of

infectious disease or CNS disorder indications was no longer observed in this sensitivity analysis. A new variable, RCT duration, whose effect was not observed in the base case model, was found to have a significant effect in this sensitivity analysis. The mean duration of RCTs included in the appraisal significantly impacted on HAS decision-making: a unit increase in trial duration decreased the odds of an ASMR V (p=0.009), and also decreased the odds of an ASMR III-IV, but this was borderline statistically significant (p=0.123). National mean drug expenditure per patient was also found to significantly impact on HAS decision-making: an increase in drug expenditure was associated with an increase in the odds of an ASMR III-IV or V, although it was only significant in the latter case (p=0.003).

The third sensitivity analysis examined the impact of a different use of ASMR ratings within the regression analysis. In the base case analysis, the ASMR ratings are grouped as follows: ASMR I-II, ASMR III-IV and ASMR V. An alternative grouping is proposed in which ASMR I-III are grouped together, followed by ASMR IV and V assessed separately. This is related to the fact that the pricing legislation in France (Social Security Code, article L. 162-16-4) specifies that the price of a technology is determined primarily by the ASMR granted to the technology, as well as price of comparator technologies, and expected volume of patients indicated for the technology. It is also specified that only those technologies that bring incremental medical value (i.e. ASMR I-IV) or a reduction in the cost of treatment can be added to the reimbursement list (Article R.163-5 du code de la sécurité sociale). Therefore, legislation specifies that those medications with an ASMR of V cannot be inserted in the reimbursement list unless they are associated with lower treatment costs. Finally, the price of medications considered to be innovative (ASMR I-III) are 'fixed' through a reference pricing mechanism, such that the price of technologies with an ASMR I-III in France cannot be lower than the price in the reference European countries. It is due to these particularities that it was considered relevant to run a sensitivity analysis in which technologies with ASMR I-III were grouped together.

The resulting model included seven variables with an accompanying pseudo R-squared value of 0.20, similar to that obtained in the base case model (0.19). With the exception of the impact of technologies for infectious diseases, all other variables, found to be significant in the base-case analyses maintained their significant impact on HAS coverage decisions even when the ASMR categorization was altered. This included the

demonstration of superiority in the clinical trial, the prevalence of the target population (and lack of such information), orphan designation, treatment for CNS disorders and musculoskeletal/joint diseases, and pharmaceutical expenditure. In the majority of variables the size and direction of the impact on the outcome variable was similar between the two types of categorization. Two exceptions are noted for two variables that were no longer statistically significant in the ASMR IV arm of the model. A unit increase in the prevalence of the target population increased the odds of an ASMR IV but was no longer statistically significant (p=0.478). Similarly, orphan drug designation no longer had a statistically significant impact on the odds of an ASMR IV (p=0.174).

**Table 7.8 Multivariate analysis of HAS coverage decisions 2004-2009: sensivitiy analysis using an alternative categorisation of the outcome variable (ASMR I-III, vs. ASMR IV or ASMR V) (n=315)**

| ASMR IV | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -0.54 | 0.139 | -1.26 | 0.18 |
| Disease prevalence | 0.00000019 | 0.478 | -0.00000034 | 0.00000073 |
| Orphan designation status | -0.73 | 0.174 | -1.78 | 0.32 |
| Central nervous system | 2.67 | <0.001 | 1.51 | 3.84 |
| Musculoskeletal and joint diseases | -1.74 | 0.030 | -3.31 | -0.16 |
| Healthcare expenditure on pharmaceuticals | 0.025 | 0.002 | 0.009 | 0.041 |
| Missing data on prevalence of disease | -1.36 | 0.025 | -2.54 | -0.17 |
| Constant | -13.35 | 0.002 | -21.67 | -5.03 |
| ASMR V | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -1.26 | <0.001 | -1.89 | -0.62 |
| Disease prevalence | 0.00000055 | 0.010 | 0.00000014 | 0.00000097 |
| Orphan designation status | -2.94 | <0.001 | -4.51 | -1.37 |
| Central nervous system | 2.27 | <0.001 | 1.13 | 3.41 |
| Musculoskeletal and joint diseases | -2.01 | 0.001 | -3.19 | -0.83 |
| Healthcare expenditure on pharmaceuticals | 0.028 | <0.001 | 0.014 | 0.042 |
| Missing data on prevalence of disease | -1.66 | 0.001 | -2.60 | -0.72 |
| Constant | -13.61 | <0.001 | -20.72 | -6.50 |

Note: Technologies with ASMR I - III are the reference case. Multinomial logistic regression, pseudo R-squared: 0.20.

In the fourth sensitivity analysis, ordinality of the outcome variable was assumed. This is in contrast with the base case analysis, where ordinality was not assumed and multinomial logistic regression was used. In this sensitivity analysis ordinal logistic regression was used. The detailed results of this analysis are provided in Appendix G, and show that assuming the ordinality assumption and modelling using a five-category outcome variable does not increase the ability for the model to explain a larger percentage of HAS decision-making, suggesting that a multinomial approach using a 3-category variable may be appropriate. The results however, of this sensitivity analysis, reveal more detail in how the factors behave within the different levels of ASMR.

## 7.4 Discussion

The overall objective of this chapter was to examine the factors that influence decisions made by the HAS to define the incremental medical value for technologies through the ASMR rating system which ranks technologies with either an ASMR I, representing the highest medical value, or ASMR V, representing no incremental medical value. In line with the hypothesised drivers of HTA decision-making highlighted in Chapter 3, and in light of evidence review presented in Chapter 2, a wide range of explanatory variables were included in the analysis, reflecting the clinical and disease characteristics of the technology under appraisal, the appraisal process itself and the socio-economic context in which the HAS operates. In addition to the general aims of the research, specific hypotheses relevant for HAS decision-making were explored and are discussed below. Unfortunately, it was not possible to incorporate insights from HAS representatives in appraising the results of this research.

### *7.4.1 Pattern of HAS decision-making*

In exploring the factors driving coverage decisions made by the HAS, 315 technology appraisals were reviewed. A primary result is that 3% of decisions were awarded an ASMR I, meaning that the technology was considered to bring highly significant incremental medical benefit and 15% of decisions were awarded an ASMR II, in instances where the commission considered that the technology would bring significant medical improvement. The majority of decisions (44%) concluded that there was no medical improvement associated with the technology (ASMR V). When HAS coverage decision were modelled, nine clinical, disease and socio-economic variables appeared to have a statistically significant impact on the odds of ASMR III-IV or ASMR V relative to ASMR I-II. No economic variables were included as HAS does not include economic criteria in its appraisal process.

### *7.4.2 The impact of clinical evidence and disease characteristics on HAS decision-making*

Given that the HAS process focuses primarily on the demonstration of incremental medical value, it was hypothesised that variables which captured the quality of the clinical evidence provided would significantly impact on HAS decision-making. Clinical and disease characteristic variables played an important role in HAS decision-making in this model. The results support the hypothesis that technologies achieving

ASMR ratings I-II are accompanied by higher quality of evidence than those technologies achieving ASMRs of III or below. Technologies that demonstrated clinical superiority increased the log odds of an ASMR I-II. Within the HAS model, as in the previous models, it was observed that the same factor could impact differently on the odds of a restriction or on the odds of a non-recommendation relative to a recommendation. For instance, while a unit increase in pharmaceutical expenditure per patient per annum decreased the log odds of an ASMR V, its effect on the odds of an ASMR III-IV was the opposite, although not statistically significant. Demonstration of clinical superiority decreased the odds of both ASMR III-IV and ASMR V. This could be seen to reflect the focus of the HAS appraisal process on the incremental benefit associated with the technology, and demonstration of clinical superiority being a key piece of evidence to substantiate the presence of incremental medical benefit.

Counter-intuitive to the hypothesis presented, the results suggest that in HAS decision-making the demonstration of superiority outranks the nature of that superiority (whether demonstrated versus placebo or an active treatment). Specifically, examination of the dataset showed that for those technologies which demonstrated clinical superiority, this superiority was demonstrated only 25% of the time versus an active comparator. For those technologies that did not demonstrate clinical superiority, 68% were compared to active comparators. It would appear that the criteria prioritized by HAS in its assessment of medical value matches closely with that used by regulatory agencies which also accept, and indeed in several circumstances recommend, comparison to placebo as an appropriate means of demonstrating clinical benefit.

When examining disease characteristics and their impact on HAS decision-making, an unexpected result was observed related to the impact of the prevalence of disease on ASMR ratings. When information on the size of the target population was lacking, this increased the odds of an ASMR I-II. This result was unexpected in that it had been hypothesized that lack of evidence should, in principle, reduce the odds of a high ASMR rating. However, upon further examination of the sample of technologies concerned (n=39), it was observed that amongst these technologies, 62% were being evaluated to renew their previously obtained reimbursement status. These technologies, by default, were therefore technologies which were already reimbursed five years previously, and which needed to renew their reimbursement status. It was likely therefore, that in these particular submissions, reference was made to previous target

population estimates, without reporting new figures. This could therefore explain the impact of non-reporting of epidemiology information within the HAS multivariate model.

It was hypothesised that ASMR ratings reflected the nature of the disease for which the technology was indicated. The analysis of HAS decision-making found that technologies that had an orphan designation decreased the log odds of an ASMR III-IV or V relative to an ASMR I-II. This could be explained by the fact that orphan designated technologies tend to be rare and indicated for diseases with no therapeutic alternatives. When the features of orphan technologies were further examined, they were found to be characterised by small patient populations (mean estimated target population for orphan technologies was 1,356 patients vs. 570,408 patients for non-orphan technologies). In addition, orphan technologies were indicated for diseases with low availability of alternative therapy (38% of cases alternative technologies were available) and therefore correspondingly higher level of clinical need for treatment. For non-orphan technologies, in 80% of cases alternative therapies were available within the French healthcare system. Orphan designation increased the odds of high ASMR ratings despite the fact that, compared to non-orphan technologies, orphan technologies on average were supported by fewer RCTs (1.5 vs. 2.8), had mean shorter trial duration (29 weeks vs. 53 weeks), were supported by smaller trials with, on average, fewer patients (255 patients vs. 1,228 patients), and had a higher proportion of instances in which an active comparator within the clinical trial was not available (34% vs. 16%). This evidence would support the hypothesis that therapeutic areas where unmet medical needs are high impact significantly on HAS decision-making. In addition, the evidence generated in this analysis suggest that the HAS may be willing to place more emphasis on the potential for the technology to fill a specific clinical need, at the expense of the quality of clinical evidence.

In addition to the effect of orphan designation on HAS decision-making, the impact of specific disease areas on coverage decisions was also examined and found to be important within HAS decision-making – providing additional evidence to support the hypothesis that ASMR ratings reflect the nature of the disease for which the technology is indicated. Technologies that had a license for the treatment of musculoskeletal and joint diseases decreased the log odds of an ASMR III-IV or V relative to ASMR I-II. In contrast, indications for the treatment of CNS disorders and infectious diseases

increased the probability of an ASMR III-IV and ASMR V relative to ASMR I-II. This could be explained by the fact that a higher proportion of technologies indicated for the treatment of musculoskeletal and joint diseases were supported by clinical trials with active comparator arms (55%). Use of active comparators in clinical trials confers more useful evidence to ascertain the incremental clinical benefit of a technology to standard of care. Placebo controlled trials are less useful in that they provide evidence of incremental benefit of a technology relative to a comparator which does not exist in clinical practice (i.e. placebo is not used to treat patients). In contrast, technologies for infectious disease or CNS disorders had a lower proportion of studies with active comparators (36% and 22%, respectively). Overall, the multivariate model emphasized the role of key clinical and disease criteria on ASMR ratings from HAS.

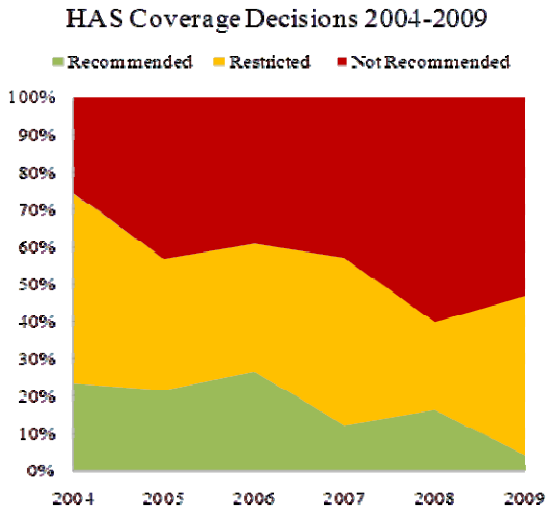
### *7.4.3 The impact of socio-economic context on HAS decision-making*

It was hypothesised that the political context in particular would impact significantly on HAS decision-making, and that this could be measured by capturing information on whether the HAS decision took place during a government election. In descriptive analyses of HAS decision-making, it was noted that there was a higher proportion of ASMR I-IV ratings than ASMR V ratings made during an election year, which was statistically significant. However, in the multivariate analysis the effect of election on HAS decision-making was no longer significant. Additional variables capturing information about the socio-economic context within which the HAS operates were also analysed. In the multivariate analysis, pharmaceutical expenditure was found to have a statistically significant impact on the log odds of both ASMR III-IV and V relative to ASMR I-II. Specifically, a unit increase in pharmaceutical expenditure, in this case national average per patient expenditure per year, appeared to increase the odds of an ASMR III-IV or ASMR V, albeit statistically significant only in the latter. This would seem counter-intuitive in that the increase in pharmaceutical spending would generally suggest an increase in the available budget for reimbursed technologies. However, a possible explanation is perhaps that an observed increase in pharmaceutical expenditure may have triggered more stringent assessment of incremental medical value. This is conjecture that cannot be further examined with this dataset, and highlights that caution is needed in the interpretation of the role of socio-economic factors such as national pharmaceutical expenditure as there are numerous unmeasured factors that could be associated with this particular variable (e.g. overall trends in GDP, change in treatment

algorithms, healthcare system approach, physician and patient behaviour, industrial policy, and increases in marketing authorisations for pharmaceuticals).

It was hypothesised that the pattern of HAS decision-making has changed over time. The descriptive analyses supported the view that outcome groups varied significantly in the mean year of appraisal. Figure 7.1 below describes the pattern of HAS decision-making over time suggesting a decrease over time in the proportion of technologies awarded an ASMR I-II and an increase in technologies receiving an ASMR V. The effect of time was however not found to have a strong effect in the multivariate analyses when taking other explanatory variables into account.

**Figure 7.1 Pattern of HAS coverage decisions 2004-2009**




### *7.4.4 Limitations*

When examining the results of the multivariate analyses, there are several limitations that need to be taken into account. An important limitation of this analysis is that it was not possible to schedule an interview with a representative of the HAS to discuss the internal validity of the model results. An additional limitation of this analysis is that HAS, compared with other agencies like NICE, provides relatively limited information in the public domain on the evidence reviewed and considered in its decision-making process. In general, HAS reports made publicly available provide concise summaries of key issues in a summarised format that do not document details on the various clinical considerations or disease characteristics which were considered. The lack of detail in reporting did lead to instances of non-reporting of information of interest for this research. The lack of data linked to this reporting style was managed by using imputation techniques in the multivariate analysis. The implications of using such

techniques, versus restricting the analysis to complete observations were assessed in a sub-analysis in which the multivariate analysis was conducted on the sample of coverage decisions for which the data was complete. The results of these sensitivity analysis showed that limiting the analysis to complete observations, may lead to bias in the coverage decisions included in the analysis. Despite this potential for bias, the sensitivity analysis regression model confirmed the important impact of several variables including: the demonstration of superiority in the clinical trial which increased the probability of an ASMR I-II, the prevalence of the target population, orphan designation, and treatment of musculoskeletal/joint diseases. The impact of infectious disease or CNS disorder indications was no longer observed in this sensitivity analysis.

Another important factor to take into account when examining HAS coverage decisions is that the sample of appraisals included in the analysis was in fact a subset of the total pool of appraisals conducted by HAS. This sub-sample of technologies used for analyses corresponded to those technologies that had been appraised by NICE and the SMC in 2004-2009. All the HAS recommendations linked to these technologies were extracted for review. The rationale for this approach was due to the fact that HAS has numerous responsibilities, one of which is the provision of advice on new technologies available for patients. In total, the Transparency Commission issued more than 2600 recommendations related to medications in 2004-2009 (HAS 2009). Given the resource constraints available, it was not possible to review all 2600 recommendations to identify those of relevance for this research (i.e. not all recommendations provide ASMR, some recommendations are related to new mode of administration, new safety information or a re-review of technologies licensed prior to 2004). The benefit of this approach was that it increased the opportunity for comparability across agencies by collecting information on a common list of compounds, and secondly it facilitated the streamlining of data extraction to those appraisals relevant for the research question.

A further limitation associated with the HAS analyses lies in the fact that it opens a window into the understanding of the factors driving HAS assessment of the degree to which technologies bring incremental medical benefit to the French population. It does not however, provide information on the degree to which the French healthcare system is willing to pay for such incremental benefit. This is directly linked to the role of HAS which is focused on evaluating the medical benefit of the technology. The output of

this assessment is then used by a separate organisation (CEPS) to negotiate a final price for the technology in question. Thus, understanding of the HAS decision-making may not provide a full perspective on coverage decisions within France and how public funding is allocated to pharmaceuticals. On the other hand, it is recognised within French legislation that the degree of medical benefit (as defined by the ASMR) directly impacts on price. Technologies with no incremental benefit are not included on the reimbursement list unless they are discounted to the available treatments already reimbursed by the system. Technologies with an ASMR IV can obtain, at a maximum, a parity price to the already reimbursed comparator. It is only technologies with a ASMR I-III rating that can aspire to potential premium prices. The pricing negotiations and discounted prices are not available in the public domain, thus preventing the inclusion of an economic component in the HAS analyses. Thus, while the multivariate analyses of HAS decisions presented here cannot directly examine the economic value that the French system attaches to particular degrees of medical benefit, it does provide an indirect view, by examining the factors that drive HAS allocation of ASMR ratings to the technologies it assesses.

In summary, the overall objective of this chapter was to examine the factors that influence coverage decisions made by HAS during the period 2004-2009 with a focus on research hypotheses specific to HAS decision-making. The results suggest that the variability in coverage decisions observed can be explained by a combination of clinical, disease, and socio-economic criteria. Strong evidence was provided to support the hypothesis that technologies supported by high quality evidence to support the incremental medical value of the technology have a decreased log odds of obtaining ASMR III or less, relative to ASMR I-II. Evidence was also shown which supported the hypothesis that ASMR ratings reflect the nature of the disease for which the technology is indicated – technologies indicated within therapeutic areas characterised by high levels of unmet need (e.g. rare disease and/or cancer) had decreased log odds of obtaining ASMR III or less, relative to ASMR I-II. The socio-economic context within which the HAS operates was also shown to play a role. Pharmaceutical expenditure was shown to have a significant effect HAS decision-making in the multivariate analysis, although this was not in the direction anticipated - increasing pharmaceutical expenditure appeared to increase the odds of an ASMR III-IV or ASMR V relative to an ASMR I-II. Descriptive analyses suggested that HAS decision-making has changed over time, and that it may be influenced by the presence of governmental elections.

Neither time nor the presence of an election were shown to impact significantly in multivariate analyses of HAS decision-making.

## 7.5 References

Bellanger M. M., Cherilova V., and Paris, V. 2005. The health basket in France, *Eur J Health Econ*, 6 (suppl. 1): 24-29.

Centre national de la recherche scientifique (CNRS). 2004. Plans pluriannuels d'action en santé publique et programmes nationaux, France. http://chronisante.inist.fr/spip.php?article177. Viewed 14January 2010

Econ-Sante` France. 2010. Dépense de médicaments. http://www.ecosante.fr/affmulti.php?base=FRAN&valeur=&langh=FRA&langs=FRA&sessionid=&TabType=0&valeur=. Viewed 3 May 2010

European Medicines Agency. 2011. European public assessment reports. http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&murl=menus/medicines/medicines.jsp&mid=WC0b01ac058001d125. Viewed between 1st of June 2009-30th March 2010.

Eurostat. 2010. Total population statistics. http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&language=en&pcode=tps00001&tableSelection=1&footnotes=yes&labeling=labels&plugin=1. Viewed 3 September 2010.

Haute Autorité de Santé . 2004. Décret no 2004-1398 du 23 décembre 2004 relatif aux médicaments remboursables par l'assurance maladie et modifiant le code de la sécurité sociale (deuxième partie : Décrets en Conseil d'Etat). http://www.has-sante.fr/portail/upload/docs/application/pdf/ctdecret2004_1398__23_12_2004_2006_09_06__17_52_20_458.pdf

———. 2006. Commission de la Transparence Avis 18 octobre 2006 HUMIRA. http://www.has-sante.fr/portail/upload/docs/application/pdf/humira.pdf. Viewed 02 January 2010.

———. 2009. Rechercher un avis. http://www.has-sante.fr/portail/jcms/c_5267/actes-medicaments-dispositifs-medicaux?catName=true&replaceFileDoc=false&searchInFiles=true&portlet=c_63468&cid=c_5267&text=&dateType=pdate&dateSince=&dateSince_user=&dateSince_unit=86400000&beginDay=1&beginMonth=0&beginYear=2004&en

dDay=31&endMonth=5&endYear=2009&typesf=generated.AVISMedicament&opSearch=Lancer+la+recherche). Viewed 2 July 2009.

———. 2011. Commission de la transparence. http://www.has-sante.fr/portail/jcms/c_419483/commission-de-la-transparence. Viewed on September 22nd 2011.

Institut de recherche et documentation en économie de la santé (IRDES). 2009. Depense de Sante- Consommation médicale totale 2009. http://www.irdes.fr/EspaceEnseignement/ChiffresGraphiques/Cadrage/DepensesSante/ConsoMedicaleTotale.htm

Joint Formulary Committee. 2010. British National Formulary. 60 ed. London: British Medical Association and Royal Pharmaceutical Society. http://bnf.org/bnf/index.htm

Le ministre de l'intérieur, de l'outre-mer, des collectivités territoriales et de l'immigration. 2004. Resultats de l'election Regionales 2004. France entière (résultats officiels). http://www.interieur.gouv.fr/sections/a_votre_service/resultats-elections/reg2004/index.htmlViewed 03 June 2009.

———. 2007. Resultats de l'election Presidentielle dimanche 6 Mai 2007, France entière (résultats officiels). http://www.interieur.gouv.fr/sections/a_votre_service/resultats-elections/PR2007/FE.html . Viewed 3 June 2009.

OECD. 2009. "OECD Health Data - Version: November 09". http://www.ecosante.org/index2.php?base=OCDE&langs=ENG&langh=ENG. Viewed 22 June 2010.

Secretariat General de la Commission de la Transparence. 2005. Rapport d'activite de la commission de la transparence 2004. http://www.has-sante.fr/portail/upload/docs/application/pdf/ctrapact2004.pdf

Sandier, S., Paris, V., and Polton, D. 2004. Health Care Systems in Transition – France 2004. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies, 2004.

Sorenson, C., Drummond, M., and Kanavos, P. 2008. Ensuring value for Money in Health Care: the role of HTA in the European Union. Cornwall: World Health Organization 2008, on behalf of the European Observatory on Health Systems and Policies.

Zentner, A., M. Velasco-Garrido, and R. Busse. 2005. Methods for the comparative evaluation of pharmaceuticals. *GMS Health Technol Assess* 1:Doc09.

# 8 Empirical analysis of a pooled dataset of NICE, SMC, CVZ and HAS coverage decisions

Having examined the role of clinical, economic and socio-economic factors within individual HTA bodies, this chapter examines how the same factors behave within a pooled sample of technologies appraised by different HTA bodies. This pooled analysis is entered into cautiously, with the realisation that comparative analyses of HTA coverage decisions must face significant challenges; however, it was felt appropriate to pursue such an analysis while recognising and addressing to some extent the identified challenges.

This chapter provides an empirical analysis of the pooled dataset. Firstly, the methods for the pooled analysis are summarised, including an overview of the sample used, as well as the descriptive and multivariate methods adopted. Descriptive statistics are calculated for the pooled sample and the results discussed. Subsequently, multivariate analyses are performed, accompanied by a series of sensitivity analyses. The results of these analyses are reported and explored, and limitations considered when formulating concluding remarks about the empirical analyses performed here.

## 8.1 Methods

The objective of this analysis is examine the impact of a range of evidence, process and socio-economic explanatory variables on coverage decisions made in that could help explain coverage decision-making across HTA bodies. An important objective of the pooled analysis was to describe the characteristics of the pooled data set and identify differences between HTA bodies the evidence considered, the process through which HTA bodies appraised and the socio-economic context in which the appraisals took place. In addition, the particular hypotheses relevant for the pooled analyses are highlighted in Box 8.1.

Box. 8.1 In light of the discussions outlined in Chapter 2, specific pooled analysis research objectives were to test whether:

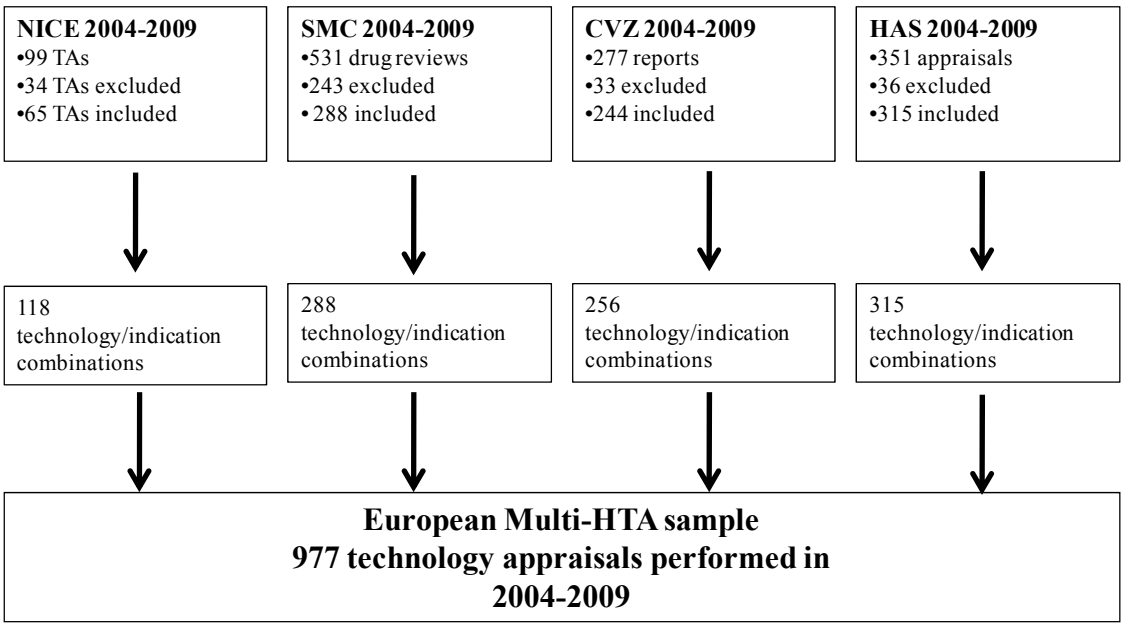
- An "HTA body effect" is observed on the odds of recommendation, restriction or non-recommendation while adjusting for a range of confounding factors.
- Among those HTA bodies that consider the ICER, the effect of the ICER is similar to that observed in the individual HTA analyses
- Process and socio-economic context variables play an important role in explaining coverage decision-making within a pooled sample of analysis

Building on from the methods described in Chapter 3, this section describes the methods used to select the sample for analysis, the outcome variable and explanatory variables considered, and the statistical techniques adopted.

### *8.1.1 Sample*

Guidance issued by NICE, SMC, CVZ and HAS in 2004-2009 was combined to create the sample for this analysis. The sample for each individual HTA body was described in Chapters 4-7. The merging of the data from the individual HTA bodies was facilitated by the fact that a similar data extraction approach was utilised and a similar set of variables extracted across the agencies. Figure 8.1 outlines the procedure followed to arrive at the pooled sample. In order to allow for effective pooling of the data, a standardised approach to defining the outcome variable was needed. This is described further below.

**Figure 8.1 Flowchart of Pooled Sample including technology appraisals performed by NICE, SMC, CVZ and HAS during 2004-2009**




### *8.1.2 Variables*

Outcome Variable

In order to be able to facilitate the analysis of the factors driving decision-making across HTA bodies, it was necessary to standardise the way the outcome variable was defined. Each agency has its own method for defining and thinking about coverage decisions. However, there are similarities in the types of coverage decisions made, which have been capitalised on to arrive at a series of 'rules' on how to define and classify coverage

decisions by NICE, SMC, CVZ and HAS to allow for comparison and pooling (Table 8.1).

The analysis is based on using a three category outcome variable where the new technology can be:

- recommended for routine use
- recommended for restricted use

or

- not recommended

HAS represents a specific challenge, in that the ASMR rating reflects the incremental value associated with a technology and impacts on the willingness of the healthcare system to approve increased funding for technologies that achieve high levels of ASMR (I or II) or restrict funding for those technologies with a low ASMR (V). The ASMR does not represent the final funding decision; rather, a separate committee, the CEPS, has this responsibility as it is the entity responsible for finalising the price and volume agreements for technologies (See Chapter 7). While recognising that the ASMR represents a different form of reimbursement decision than the coverage decisions made by SMC, NICE and CVZ, to allow for pooling a classification of ASMR ratings was adopted (described in Table 8.1). This was felt to be appropriate given that the price of the technology, and hence the funding of the technology, is primarily driven by the ASMR rating. An alternative classification was tested in a sensitivity analysis of HAS decision-making (Chapter 7) and found similar results to the base case analysis.

Recommended technologies were defined as those technologies where full coverage was granted for the totality of the licensed population. For NICE guidance, where a recommendation was made for a technology to be used in a population identical to its licensed indication, it was considered 'recommended' (See also Chapter 4). For the CVZ, where the decision was to place the technology in the 'basis paket' (List 1A or 1B), or listed in the expensive drug list (*Duregeneesmiddelen Beleidsregel*) without any restriction or patient co-payment, the technology was considered to be recommended (See also Chapter 6). In France, the ASMR was used to classify outcomes. Recommended technologies in this analysis were considered to be those with an ASMR I- II, representing technologies with important incremental medical value relative to standard of care (See also Chapter 7).

**Table 8.1 Classification of coverage decisions into a 3-category outcome variable: definitions per HTA body**

| HTA body | Recommended | Restricted | Not Recommended |
|---|---|---|---|
| **NICE** | Full coverage was granted for the totality of the licensed population | A sub-population of the licensed indication and/or with restrictions in terms of acquisition cost or utilization (e.g. monitoring or specialist use required) | "Not recommended" stated in section 1 of guidance |
| **SMC** | If word "recommended" used in summary statement | If word "restricted" was used in summary statement | If words "not recommended" were used in summary statement |
| **CVZ** | If technology was placed in reimbursement lists 1A or 1B, or the expensive drug list | If technology placed in List 2, or if patient co-payment is necessary to access medication | If the technology was not included in any reimbursement list |
| **HAS** | ASMR I-II | ASMR III-IV | ASMR V |

Restricted technologies were defined as those technologies where coverage was granted for a sub-population of the licensed indication and/or with restrictions in terms of acquisition cost or utilization (e.g. monitoring or specialist use required) (Raftery 2006). For NICE and the SMC, a coverage decision was considered to be a restriction if it was recommended for use in a sub-population of its licensed indication; in a second line or higher line of therapy; required monitoring, lowest acquisition cost or prescription by a specialist. For the CVZ, where the decision was to place the technology in the 'basis paket', but only for use in a sub-population or with a patient co-payment, this technology was considered as restricted. With regard to HAS, technologies with ASMR III-IV representing modest or minor medical value are associated with lower price levels (and hence funding) than technologies with ASMR I-II. These were considered to be restricted technologies.

Not recommended technologies were those for which no coverage was granted. A medication was considered to be not recommended for use by NICE or SMC guidance if the words "not recommended" were stated in the guidance/report. Within CVZ decisions, the technology was considered to be not recommended when it was stated to be not recommended in the CVZ 'advies' statement and was not included in any reimbursement list. With regard to HAS, technologies with an ASMR V offer no incremental benefit versus the comparators, and as per legislation cannot be included on the reimbursement list. An ASMR V was therefore considered to be 'not recommended'. It should be noted however, that for HAS, technologies with an ASMR V can obtain reimbursement from the healthcare system, but only if associated with

cost-savings. This is different from the CVZ where technologies not recommended for funding are excluded from the reimbursement list. Other research conducted on the ASMR has made a similar assumption whereby technologies with ASMR V were considered to represent non-recommendation (Kanavos et al. 2010). While for the purposes of this analysis it was felt appropriate to consider ASMR V as a non-recommendation, it is important to bear in mind that the implications of an ASMR V may not be the same as the implications of non recommendation from the SMC or NICE, for example.

Explanatory Variables

The variables included in this cross HTA body comparative analysis are consistent with the variables considered in the single HTA body analyses presented in previous chapters (4-7). The selection of variables is directly linked to the analytical framework and the variables included ranged from clinical trial and disease characteristics to economic cost-utility model information and from appraisal process characteristics to the socio-economic context in which the appraisal was performed. Across the four HTA bodies, in addition to a core set of variables, a series of variables were collected specifically for each HTA body to reflect the nature of the appraisal process as accurately as possible in the individual analyses (e.g. use of MTA vs. STA by NICE or information on the reason for the appraisal request within the HAS ). However, these variables were not included as explanatory variables in the pooled sample to allow for a more homogeneous platform upon which to compare HTA bodies using a quantitative approach. The impact of these variables was assessed in Chapters 4-7 where individual HTA multivariate analyses were run.

In addition to the core set of variables, a new variable was introduced specifically for this pooled analysis which aimed to capture the role of the HTA body itself on the impact of coverage decisions. The aim was to establish the effect of the HTA body on the odds of restriction or non-recommendation, while holding other factors constant. The statistical methods adopted in the multivariate analysis are described further below.

## 8.2 Statistical Analyses

The methods for the descriptive statistics and multivariate analyses were described in Chapter 3. Descriptive statistics were calculated for the pooled sample of NICE, SMC, CVZ and HAS. Following a descriptive analysis of the dataset, a multinomial logit

regression was modelled, according to the model specification process described in Chapter 3. The objective of this analysis was to obtain a parsimonious model that best reflected the main drivers of decision-making.

Among the challenges identified for performing a pooled analysis across HTA bodies (Chapter 3, section 3.1.5, Chapter 8, Section 8.1-8.2), a key challenge encountered was the fact that variation was observed between HTA bodies in the factors relevant to their decision-making process. The most obvious difference is linked to the HAS: the variables related to the economic characteristics of the technology under appraisal are not available, as economic considerations are not a formal criteria in the HAS appraisal process. To address this challenge, two options were examined in multivariate analyses: 1) to include all four HTA bodies in the pooled analysis accompanied by fixed effects, but to exclude variables that are not common across the four (in particular ICER-related variables); 2) exclude HTA bodies that do not consider the cost-effectiveness of technologies to avoid imputation of information that was not formally considered (Table 8.2).

An additional key challenge is the fact that there is heterogeneity between the four HTA bodies in the criteria used to select technologies for appraisal. While the selection of technologies for appraisal by NICE are related to consultation with the Department of Health, the SMC, HAS and CVZ operate on a system where all new technologies and new indications are appraised systematically. NICE, HAS and CVZ re-review a sub-set of the technologies after a specified time period, but this is not the case for SMC. This leads to a situation where the role of a specific factor and the pattern of coverage decisions observed may be confounded by the method and timing with which technologies are selected for appraisal. In the base case analysis, heterogeneity between HTA bodies in the technologies appraised is adjusted for by including BNF disease categories in the model. This was thought to be a valid approach for two reasons. Firstly, it was not possible to adjust for the technology itself, as there were not enough observations for each of the 348 technologies appraised to implement such a strategy. Secondly, it was assumed that differences in technologies appraised between HTA bodies would, in many cases, correspond with differences between HTA bodies in the disease areas appraised. It was therefore felt that by including disease category descriptors (namely BNF categories) in the model would increase the homogeneity of the sample for analysis. In addition, to examine this issue further, a sub-set of

technologies which were appraised by all four HTA bodies was used as the basis for a sensitivity analysis. Similarly, a sensitivity analysis was also performed restricting the analysis to technologies indicated for the treatment of cancer diseases (Table 8.2). The analysis was performed using the STATA data analysis software (Intercooled (IC) Stata version 10.1).

**Table 8.2 Multivariate analyses performed on pooled data set – base case and alternative specifications**

| | **Key attributes of analysis** | **Rationale** | **Analysis characteristics** | **Results shown** |
|---|---|---|---|---|
| Base Case1 | Analysis without economic variables | To assess whether HTA bodies differ amongst themselves in terms of the coverage decisions they make, while adjusting for a range of confounders | N=977<br><br>Included HTA bodies: NICE, SMC, CVZ, HAS<br>Excluded variables: ICER (uncertainty estimates), budget impact assessment<br>Includes fixed effect for HTA body | Ch. 8, 8.3.3, Table 8.3 |
| Base Case 2 | Analysis with economic variables excluding HAS | Explore the degree of impact of a range of evidence, process and socio-economic factors on coverage decisions.<br>Excluding HAS allows for economic variables to be explored without imputing economic values for HAS appraisals | N=662<br><br>Included HTA bodies: NICE, SMC, CVZ<br>Economic Variables included | In appendix H, Table 8.5 |
| Base Case 3 | Analysis of NICE&SMC only | Explore the degree of impact of a range of evidence, process and socio-economic factors on coverage decisions.<br>NICE & SMC frequently use cost-utility evidence as part of appraisal (unlike HAS and CVZ (less than 11% of decisions report ICER in latter) | N=406<br><br>Included HTA bodies: NICE, SMC<br>Economic Variables included | Ch. 8, 8.3.3, Table 8.4 |
| Alternative specifications | Analysis utilising binary outcome variable | To examine the effect of alternative methods for categorising coverage decisions on model outcome | Run on base case 1<br>N=977<br>Binary outcome category used (covered vs. not covered) | Ch. 8, 8.3.3, Table 8.5 |
| | Analysis including cancer therapies only | To standardise the baseline sample used in the multivariate analysis | Used base case 1<br>N= 247 | Ch. 8, 8.3.3, Table 8.6 |
| | Analysis including common set of technologies only | To standardise the baseline sample used in the multivariate analysis | Run on base case 1<br>N=192 | Ch. 8, 8.3.3, Table 8.7 |
| | Analysis with economic variables including HAS | To keep all four HTA bodies in the pooled analysis and to consider in the analysis the factors shown to have impact on individual HTA body analyses<br>Imputation used to generate values for economic indicators in HAS dataset with accompanying interaction term | N=977<br>All HTA bodies included<br>Economic variables included | In appendix H, Table 8.4 |

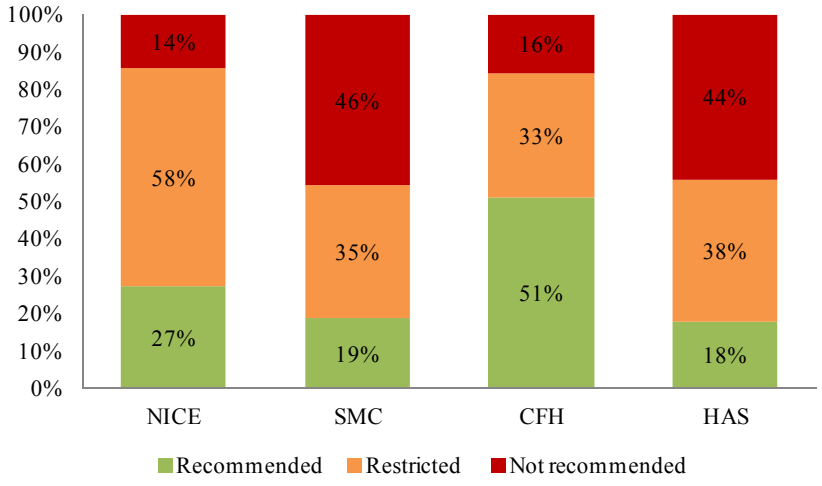
## 8.3 Results

### *8.3.1 Sample characteristics*

Outcome Variable

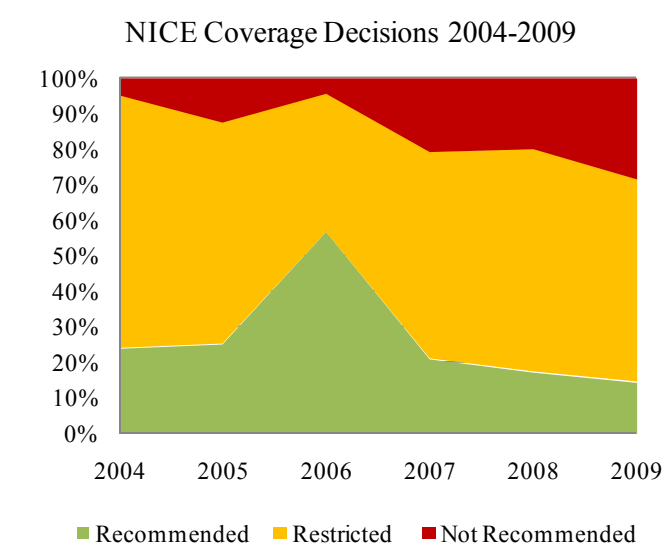
In total, 977 HTA decisions were reviewed and analysed, including 118 NICE decisions, 288 SMC decisions, 256 CVZ decisions and 315 HAS decisions made during 2004-2009. Within the pooled data set, 27% of decisions recommended funding of the technology, 39% restricted funding and 35% did not recommend funding the technology.The most common decision by NICE was to restrict funding (58%), whereas for the CVZ the most common decision was to recommend (51%). For both the SMC and HAS, the most common decision was to not recommend (46% and 44%). For NICE and CVZ, the least common coverage decision was to not recommend (14% and 16% respectively) while for SMC and HAS the least common decision was to recommend (19% and 18% respectively). It is noteworthy that the outlier in terms of recommendations, without any restrictions, was the CVZ, with 51% of its decisions to recommend funding for treatment. This compares to 27% for NICE, 19% for SMC and 18% for HAS (Figure 8.2).

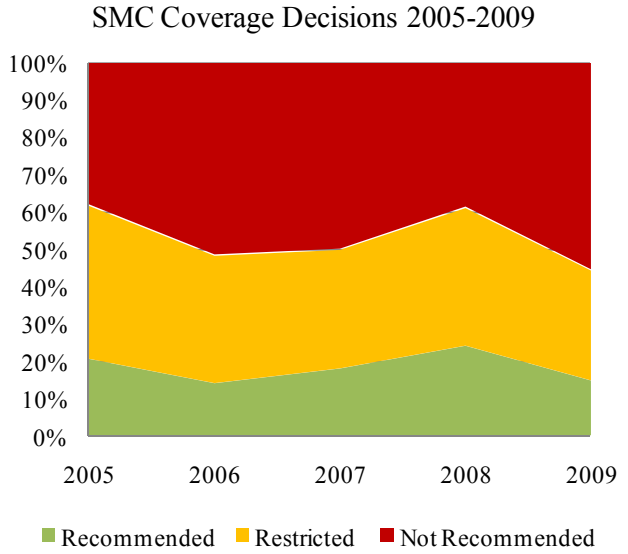
**Figure 8.2 Coverage decisions by NICE, SMC, CVZ and HAS between 2004-2009 Total sample (n=977)**

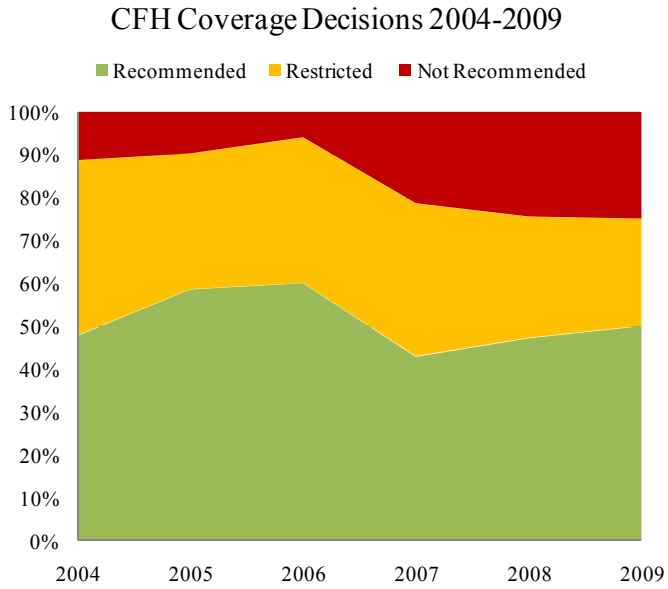


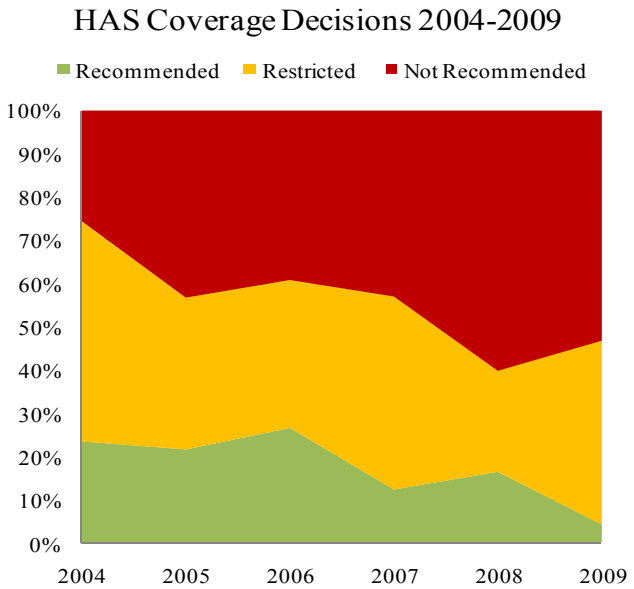

The trends in coverage decisions over time (January 2004- June 2009) for each HTA body are presented in Figure 8.3. These data suggest that within HAS and NICE there has been a decrease in the proportion of recommendations made over time and a corresponding increase in the proportion of non-recommendations. The CVZ and SMC on the other hand, appear to maintain relatively stable rates of recommendations, restrictions and non-recommendations over time.

**Figure 8.3 NICE, SMC, CVZ and HAS coverage decisions between 2004-2009 (June), by year (n=977)**



Coverage decisions across HTA bodies were also described for two sub-samples, both of which were designed to increase the homogeneity of the technologies appraised. The first sample represented the set of coverage decisions linked to a common basket of technologies appraised across all four HTA bodies (Figure 8.4). In this particular sub-sample, the pattern of coverage decisions observed in NICE, SMC and HAS datasets was similar to that observed in the total sample, while for the HAS, a lower proportion of technologies were not recommended (20% versus 44% in the total HAS sample). In the second sub-sample of appraisals that examined cancer therapies (n=247), the pattern of coverage decisions by HTA agencies was different in that recommendations appeared to be more common than when examining the pattern of coverage decisions across the total sample, with the exception of the SMC (Figure 8.5). The SMC was the only agency in which a small decrease in recommendations was observed when the analysis was restricted to cancer therapies alone (from 19% to 15%), and in general the coverage pattern for the SMC did not appear to alter when considering the total SMC pool and the cancer technology subset. Unlike the SMC, NICE recommendations increased from 27% to 36%; CVZ recommendations from 51% to 76% and HAS recommendations from 18% to 29%. While this trend needs to be examined more carefully with appropriate statistical methods, it does suggest that comparison of coverage decisions across HTA bodies needs to take into account the characteristics of the technologies included in the sample. Careful interpretation of multivariate analyses, and sensitivity analyses to test base case model specifications is needed, when attempting to compare HTA decision outcomes and factors driving those decisions.

**Figure 8.4 Technologies common to all HTA bodies- coverage decisions by NICE, SMC, CVZ and HAS between 2004-2009 (n=192)**

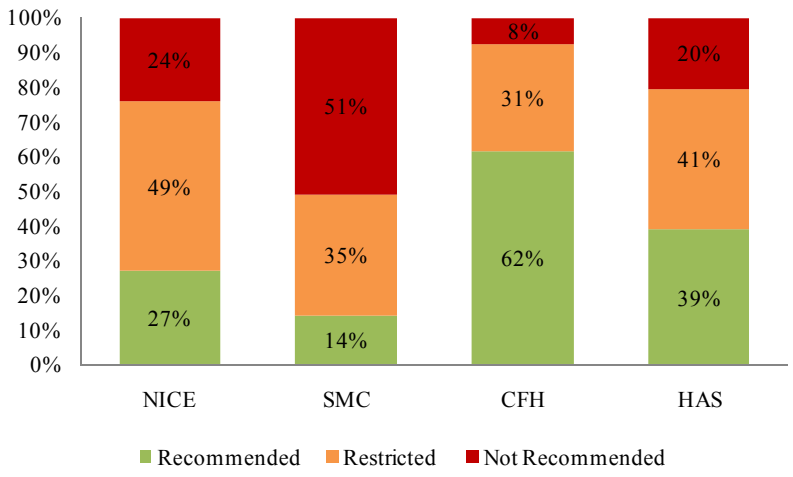
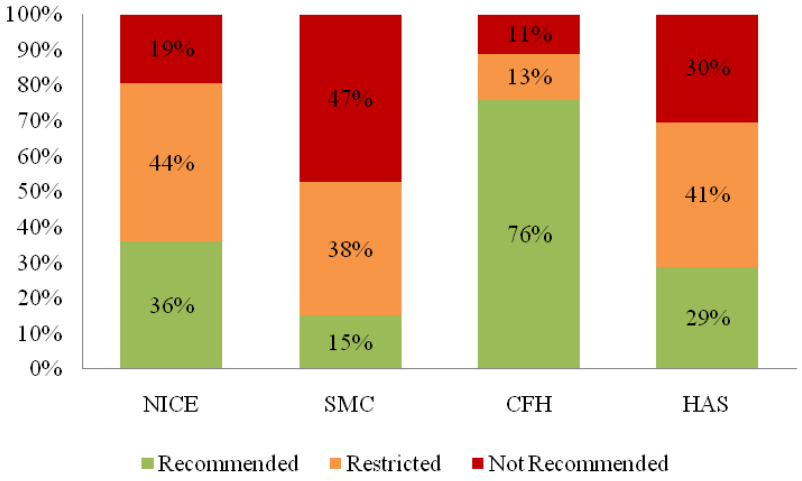


**Figure 8.5 Cancer therapies - coverage decisions by NICE, SMC, CVZ and HAS between 2004-2009 (n=247)**




### *8.3.2 Descriptive statistics*

Descriptive statistics were calculated for the total sample as well as for each decision outcome. These are summarized in Appendix H Tables 1 – 5. Results for each category of explanatory variable are provided below.

Clinical Evidence

The clinical characteristics of the technologies appraised were statistically significantly different across the coverage decision types (Appendix H Table 1) as well as across the four HTA bodies in terms of the number of RCTs considered, their size, statistical significance of RCT results, follow-up and comparator used within the trials (all variables statistically significant at $p \leq 0.01$) (Appendix H Table 2). The greatest difference was observed between NICE and the remaining three HTA bodies. The mean number of RCTs considered in SMC, CVZ and HAS decision making was 2-3 studies (compared with 7 studies for NICE), and the duration of follow-up that they considered in the appraisal ranged from 39-49 weeks compared to 76 weeks for NICE. The RCT mean sample size ranged from 830 in appraisals by the CVZ to 1249 by NICE, and differences were statistically significant. On average 42% - 52% of RCTs included active comparators (as opposed to placebo comparators), and differences between the HTA bodies were demonstrated to be statistically significant. When examining the nature of the clinical evidence across recommended, restricted and not recommended technologies, the latter tended to be supported by a fewer number of trials of shorter duration and size than recommended and restricted technologies.

Descriptive analysis of the clinical package included in the appraisal process was also conducted for each decision outcome. Within the sub-set of recommended, restricted and not recommended technologies, the pattern observed was similar to that noted in the total sample (Appendix H Tables 3-5). That is, technologies appraised by NICE tended to be supported by a higher number of trials of longer duration and larger sample size than technologies reviewed by the other three HTA bodies. The majority of technologies restricted by NICE did not demonstrate superiority in efficacy. In contrast, approximately half of SMC and HAS restricted technologies had demonstrated statistically significant superiority. However, in 40%-53% of cases superiority was demonstrated versus placebo rather than an active comparator. Amongst the appraisals by NICE, HAS and CVZ that resulted in non-recommendation, approximately 65% of the technologies did not demonstrate statistically significant superiority in their RCTs. SMC was a clear outlier in this regard, where 58% of technologies that were not recommended did demonstrate statistically significant superiority ($p<0.01$). Within the non-recommended technologies, the proportion with active comparators as opposed to placebo comparators differed between HTA bodies ranging from 21% in CVZ appraisals to 50% in HAS appraisals ($p<0.05$).

The technologies appraised also differed, across coverage type and HTA body, in terms of the diseases for which they were indicated. Across all HTA bodies, the majority of appraisals were conducted for technologies where alternative therapies were already available in the healthcare system ($p<0.01$). Within the recommended technologies, approximately 90% of recommendations by the SMC and NICE were for technologies where alternatives were available, while for the HAS and CVZ alternatives were available in approximately 70% of cases ($p<0.05$). A similar pattern was observed within the restricted and not-recommended technologies.

Orphan designated medicines made up approximately 10% of CVZ, SMC and HAS technologies appraised, while NICE considered very few orphan medicines (3 of 118 technologies) ($p<0.05$). Within the sample of recommended technologies, the proportion of recommendations for orphan medicines differed considerably between HTA bodies ($p<0.01$): 24% of HAS recommendations were orphan medicines, in contrast to 6% of SMC and 5% of CVZ recommendations. Among the sub-set of technologies that were not recommended for funding, very few were orphan medicines

within the HAS (1%), NICE (0%) and CVZ (0%) samples. In contrast, in SMC appraisals, 18% of non-recommended technologies were orphan medicines.

The proportion of technologies indicated for the treatment of musculoskeletal diseases, obstetrics and endocrine system disorders differed significantly between HTA bodies. Within the recommended technologies, a higher proportion of SMC technologies were indicated for infectious diseases than compared with other HTA bodies, while none of the recommended SMC technologies were indicated for musculoskeletal and joint diseases. Within the sub-set of restricted technologies, there was variation in the proportion of restricted technologies indicated for malignancies and musculoskeletal diseases across HTA bodies ($p<0.01$), as well as in the proportion of technologies indicated for obstetrics and gynaecology disorders. These differences across HTA bodies were also observed in the sub-set of technologies that were not recommended for public funding.

In terms of population treated, the prevalence of the diseases considered in the appraisals were for NICE 2.4 million patients, SMC 0.11 million patients, CVZ 0.94 million patients, and HAS 0.51 million patients. Amongst the recommended technologies, the prevalence of the diseases linked to recommended technologies were for NICE 0.4 million patients, SMC 0.04 million patients, CVZ 0.061 million patients and HAS 0.139 million patients. The prevalence of the diseases considered in the appraisals that resulted in restricted funding were for NICE 3.2 million patients, SMC 0.07 million patients, CVZ 0.04 million patients, and HAS 0.24 million patients ($p<0.01$). Within the subset of technologies not recommended for funding, the prevalence of the diseases considered in the appraisals were 2.3 million patients for NICE, 0.004 million patients for the SMC, 0.27 million patients for the CVZ, and HAS 0.91 million patients.

Economic Evidence

HAS does not formally consider economic criteria in its appraisal process, therefore no results for the HAS are presented in this section. The three remaining HTA bodies differed markedly in the percentage of appraisals that considered cost-utility evidence and this difference was statistically significant ($p<0.01$). Of the three HTA bodies, NICE most frequently considered cost-utility evidence, as opposed to the CVZ which rarely considered it. This pattern was observed within the total pooled sample, as well

as in the subset of recommended, restricted and not-recommended technologies. In the total sample, cost-utility evidence was considered in the majority of appraisals by NICE (95%), followed by the SMC (74%). Within the subset of recommended technologies, 100% of technologies recommended by NICE were supported by CUA, compared with 67% for the SMC, and 12% for the CVZ. Among those technologies that were not recommended for public funding, the following proportion of technologies were supported by cost-utility analyses: 88% of NICE appraised technologies, followed by the SMC (75%) and the CVZ (18%). Amongst technologies appraised by the CVZ, only a small proportion of technologies (<18%) included cost-utility analysis, despite the inclusions of cost-effectiveness criteria in its formal appraisal process in 2006.

Not only was NICE the most likely to review cost-utility evidence in its appraisals, it was also most likely to review several cost-utility models – 63% of its appraisals contained reference to more than one cost-utility model. This was not the case in either the SMC or CVZ appraisals, and this difference was statistically significant ($p<0.01$). The use of multiple cost-effectiveness models by NICE captured in the total sample was observed consistently across recommended, restricted and not recommended technology subsets.

Among the reported cost-utility results, the average cost-utility ratio across the three decision outcomes was statistically significantly different ($p<0.05$). When comparing the mean cost per QALY across HTA bodies the following ICERs were observed: £31,266 cost per QALY for NICE and £34,055 cost per QALY for the SMC. For the few appraisals that contained cost-utility analyses in CVZ appraisals, the average ICER was £30,977 per QALY. The SMC most frequently considered non-cost utility analyses – 30% of its appraisals considered such economic data, compared to NICE (23%) and CVZ (15%).

In terms of budget impact, the average budget impact estimated for the appraised technology was £701 million for NICE, compared with the £1.192 million for the SMC and £31 million for CVZ. In terms of budget impact for recommended technologies, the average budget impact estimated for the appraised technology was lower in the recommended group that in the total sample: £36 million for NICE, £1.9 million for the SMC and £5.8 million for CVZ. For the restricted technologies, the average budget impact estimated for the appraised technology was £829 million for NICE, compared

with the £1.3 million for the SMC and £66.8 million for CVZ ($p<0.01$). Mean budget impact estimates within the non-recommended subset of technologies were considerably higher than in restricted or recommended technologies, with the exception of the SMC: the average budget impact estimated for the appraised technology was £1632 million for NICE, compared with the £0.9 million for the SMC and £63 million for CVZ ($p<0.01$).

Appraisal process

Patient group submissions were commonly included as part of the evidence appraised by NICE – in 87% of appraisals compared to 42% compared to the SMC. None of the HAS submissions included evidence from patient group submissions as this is not a formal component of the appraised evidence, and the CVZ very rarely reported the inclusion of patient group submissions (4% of appraisals). These differences in use of evidence from patient group submissions was statistically significant ($p<0.0001$). A similar pattern was observed within the recommended sample of technologies: in 91% of NICE appraisals compared to 27% SMC appraisals and 2% of CVZ appraisals. In comparison with the recommended technologies, the proportion of restricted technologies supported by patient submissions was similar for NICE (84% vs. 91%), but twice as high for the SMC (45% of restricted technologies supported by patient submissions vs. 27% for recommended technologies), and 2% of CVZ appraisals. The pattern observed within the recommended technologies was similar to that observed within the not recommended technologies: 94% of NICE appraisals compared to 44% of SMC appraisals and 13% of CVZ appraisals ($p>0.01$).

Information on the number of decision-makers involved in the appraisal process was available for all four HTA bodies. For the CVZ and HAS, a fixed number of members are involved in each appraisal ($n=20$ and $n=31$ respectively), while for NICE and SMC this appeared to vary: on average 30 members in NICE appraisal committees and 25 in SMC committees. The differences observed in the size of the appraisal committees were shown to be statistically significant ($p<0.01$) and were consistently observed across decision outcomes.

The cost-effectiveness component was a formal part of the appraisal process for both NICE and the SMC, and became part of the formal process by the CVZ from 2006 onwards. Thus, for the CVZ approximately 67% of appraisals were conducted after 2006 when the cost-effectiveness component became part of the process. As seen above

however, this did not mean that all appraisals post-2006 included cost-effectiveness evidence. In terms of budget impact, NICE, SMC and CVZ included a budget impact component in their decision-making. The HAS did not have economic components within its appraisal process. The differences between HTA bodies in the use of the cost-utility component, as well as the budget impact components, in the appraisal process were statistically significant ($p<0.01$).

Of the four HTA bodies, NICE is the only HTA body that can appraise technologies individually or collectively. This led to a statistically significant difference in the average number of technologies appraised across the HTA bodies. For recommended technologies, the average number of technologies appraised by NICE was 2 compared with 1 technology for the remaining HTA bodies ($p<0.01$). Within the restricted sample, the mean number of technologies appraised by NICE was 3.4, and within the not recommended group the mean number of technologies appraised by NICE was 2.

Socio-economic context factors

Information on a selection of socio-economic factors was collected, including population size, healthcare expenditure, whether there was an election year and what the priority disease areas were for the healthcare systems in which the HTA bodies were operating. The population to which the HTA systems made their recommendations ranged from 5 million in Scotland to 63 million in France ($p<0.01$). Healthcare expenditure, as measured by percentage of gross domestic product (GDP) ranged from an average of 8% in England and Wales and Scotland to 11% in France ($p<0.01$). In terms of average expenditure on pharmaceuticals, this ranged from £175 per patient in England and Wales to £439 per person in France, and was statistically significantly different ($p<0.01$). Interestingly, the proportion of decisions made within an election year varied between HTA body ($p<0.01$): 7% of decisions by NICE were made during a governmental election year, as opposed to 40% by the SMC. In the Netherlands, 20% of CVZ decisions were made in an election year and in France, 30% of HAS decisions were made in an election year. Relatively similar proportions of appraisals were conducted for technologies related to the treatment of priority diseases.

A descriptive analysis of the socio-economic context was also conducted for each decision outcome – although in general, as the socio-economic factors were external to decision-making, while variation in these factors was observed between HTA bodies,

there was no additional variation observed between decision outcomes across HTA bodies. A larger proportion of recommendations by the CVZ, SMC and HAS were made in an election year, compared with NICE: 6% of recommendations by NICE were made during a governmental election year, as opposed to 48% by the SMC, 23% by the CVZ and 31% by the HAS ($p<0.01$). For NICE, the proportion of decisions made within an election year remained the same across coverage categories (6%-7%). For HAS there were fewer non-recommendations made in an election year (25% vs. 31% recommendations and 37% restrictions). Similar to HAS, the CVZ saw a lower proportion of non-recommendations (8%) made in an election year than recommendations (23%) and restrictions (20%).

Summary of descriptive statistics

Within the pooled sample of NICE, SMC, CVZ and HAS coverage decisions, of the explanatory variables examined, descriptive analyses suggest that the following factors may play an important role in determining coverage decision-making. For these variables (Table 8.3), statistically significant differences were observed between interventions that were recommended, restricted and not recommended ($p \leq 0.05$).

**Table 8.3 Pooled HTA sample: statistically significant variables ($p \leq 0.05$)**

| **Component 1 – Evidence characteristics** | **Variables** |
|---|---|
| Clinical Package | Number of RCTs reviewed<br>Superiority in primary endpoint<br>RCT duration of follow-up<br>Use of observational studies |
| Economic Package | Use of Multiple CE models<br>ICER<br>Probabilistic sensitivity analysis<br>Univariate uncertainty estimates |
| Disease characteristics | Prevalence<br>Disease indications<br>Therapies indicated for: infectious diseases, central nervous system, ear/nose, eye, malignant diseases, musculoskeletal/joint conditions, obstetrics, respiratory diseases |
| **Component 2 – Process characteristics** | Inclusion of patient submission/evidence<br>Number of decision-makers<br>CE included in process<br>Budget Impact Assessment included in process<br>Number of technologies appraised<br>HTA body |
| **Component 3 – Socio-economic characteristics** | Year of appraisal<br>National Population<br>GDP health expenditure<br>Pharmaceutical expenditure per patient per year |

### *8.3.3 Multivariate analysis*

Following the model specification process described in Chapter 3, three base case analyses were conducted, in accordance with Table 8.2. Base case model 1 included all four HTA bodies, but excluded economic variables as explanatory variables in the model. The model included 20 variables yielding a pseudo R-squared of 0.13, suggesting that the model explains approximately 13% of the variability in coverage decisions across the pooled sample (Table 8.4). Base case model 2 included NICE, SMC and CVZ and included economic variables in the analysis. The HAS was excluded from the analysis as it does not formally incorporate cost-effectiveness considerations in its appraisal process. In this analysis, the ICER was found not to have a statistically significant impact on the log odds of restriction versus recommendation or non-recommendation versus recommendation. When the rationale for this result was explored, attention was drawn to the low reporting of ICERs within the CVZ appraisals (11% of appraisals reported ICERs see Chapter 6, Table 6.4). This is in part driven by the fact that cost-effectiveness considerations were formally introduced in the CVZ process in 2006 and cost-effectiveness results are only reported for those technologies that are associated with incremental therapeutic benefit to patients. The results of base case 2 model are displayed in Appendix H, Table 8.5. In light of the results observed in base case model 2, a third base case analysis was conducted which included both NICE and the SMC, incorporating the ICER as an explanatory variable. The model included 19 variables yielding a pseudo R-squared of 0.16, suggesting that the model explains approximately 16% of the variability in coverage decisions across the pooled sample (Table 8.5).

With regards to the effect of clinical variables within these base case analyses, the number of RCTs, as well as their duration, choice of comparator and design, impacted significantly on the log-odds of restriction or non-recommendation versus recommendation. Specifically, if the technology demonstrated clinical superiority, the log-odds for restriction relative to recommendation decreased ($p=0.023$), as did the log-odds of non recommendation relative to recommendation ($p=0.001$), while holding all other variables constant. A unit increase in the number of RCTs, and a unit increase in their duration increased the log-odds of a non-recommendation, and had the same effect on the odds of a restriction, although this was not statistically significant. The use of an active comparator as opposed to a placebo comparator within the clinical trial was a

significant explanatory influence on both the log-odds of restriction and non recommendation.

With regard to economic variables, these were examined in base case model 3, combining appraisals by NICE and SMC in a pooled analysis (Table 8.5). A unit increase in the ICER was shown to increase the odds of restriction relative to recommendation (p=0.011) and the log odds of non-recommendation, relative to recommendation (p=0.001). These results confirmed the important role of the ICER in NICE and SMC decision-making, while adjusting for a range of confounders.

Process factors also had significant impact on coverage decisions in this pooled analysis. In base case model 1, an increase in the number of technologies appraised simultaneously exerted a increased impact on the log-odds of a restriction (p=0.002), and the log-odds of a non-recommendation (NS). In the model examining NICE and SMC (base case 3), an increase in the number of technologies significantly increased the odds of both restriction and non-recommendation relative to recommendation (p=0.001 and p=0.074, respectively). The inclusion of patient submissions and patient evidence as part of the process was linked with an increase in the log-odds of a restriction and non-recommendation, although this was statistically significant in the latter case only (p=0.008) in the base case model 1.

Of interest is the role of the HTA bodies themselves in explaining variation in coverage decisions. In base case model 1, which incorporated all four HTA bodies, when the impact of NICE, SMC and HAS on coverage decisions was examined relative to the CVZ, the results suggest that NICE and HAS assessment bodies are strongly associated with a decreased odds of a restriction or non recommendation. This can be contrasted with the effect of the SMC, which was found to statistically significantly increase the log-odds of both restriction and non-recommendation in all base case models. The impact of the HTA body was highly statistically significant across all assessments.

Socio-economic factors also contribute to explaining the variability in coverage decisions across the HTA bodies. With regard to the size of the population within the HTA body remit, a unit increase in the population size increased the odds of both restriction and non-recommendation, and both effects were statistically significant in both base case models 1 and 3.

**Table 8.4 Base case 1: BASE CASE 1 with all four HTA bodies (but no ICER, and including fixed effects for each HTA body) (n=977)**

| | Restricted vs. Recommended | | | | Not Recommended vs. Recommended | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.018 | 0.402 | -0.061 | 0.025 | -0.060 | 0.058 | -0.122 | 0.002 |
| RCT duration of follow-up | -0.001 | 0.453 | -0.0044 | 0.00198 | -0.003 | 0.112 | -0.007 | 0.001 |
| Use of active comparator in RCT | -0.535 | 0.011 | -0.949 | -0.121 | -0.932 | <0.001 | -1.383 | -0.481 |
| Clinical superiority demonstrated in RCT | -0.433 | 0.023 | -0.807 | -0.059 | -0.762 | <0.001 | -1.176 | -0.349 |
| Size of eligible patient population | 5.00E-07 | 0.093 | -0.00000134 | 0.00000010 | 0.000 | 0.633 | 0.000 | 0.000 |
| Orphan designation status | -0.311 | 0.294 | -0.891 | 0.269 | -0.749 | 0.023 | -1.397 | -0.101 |
| Patient submission included | 0.391 | 0.203 | -0.211 | 0.993 | 0.833 | 0.008 | 0.215 | 1.452 |
| Number of technologies appraised simultaneously | 0.523 | 0.002 | 0.197 | 0.849 | 0.146 | 0.546 | -0.328 | 0.620 |
| National population size | 0.0000005 | 0.024 | 0.00000007 | 0.0000009 | 0.000 | <0.001 | <0.001 | <0.0001 |
| Central nervous system | 0.213 | 0.464 | -0.357 | 0.784 | 0.381 | 0.209 | -0.213 | 0.976 |
| eye | -1.443 | 0.047 | -2.866 | -0.021 | -1.314 | 0.075 | -2.761 | 0.134 |
| Malignancy/immunosuppresion therapy | -0.532 | 0.031 | -1.016 | -0.048 | -0.457 | 0.087 | -0.981 | 0.067 |
| musculoskeletal and joint diseases | -0.200 | 0.536 | -0.833 | 0.433 | -0.958 | 0.017 | -1.747 | -0.169 |
| obstetrics, gynaecology, and urinary-tract disorders | 2.292 | 0.034 | 0.171 | 4.413 | 1.700 | 0.148 | -0.603 | 4.003 |
| Respiratory system | 0.492 | 0.511 | -0.974 | 1.958 | 1.747 | 0.012 | 0.378 | 3.116 |
| Cardiovascular disease | 0.067 | 0.828 | -0.540 | 0.675 | -0.290 | 0.39 | -0.952 | 0.372 |
| Skin | -0.011 | 0.980 | -0.856 | 0.834 | -0.302 | 0.529 | -1.242 | 0.639 |
| NICE | -18.526 | 0.027 | -34.957 | -2.096 | -37.714 | <0.001 | -55.265 | -20.163 |
| SMC | 6.666 | 0.0080 | 1.761 | 11.571 | 13.270 | <0.001 | 8.034 | 18.506 |
| HAS | -22.156 | 0.033 | -42.550 | -1.762 | -44.951 | <0.001 | -66.727 | -23.175 |
| Constant | -8.412 | 0.022 | -15.589 | -1.235 | -16.638 | <0.001 | -24.299 | -8.977 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.26.

**Table 8.5 Base-case 3: Multivariate analysis of pooled sample of NICE and SMC coverage decisions 2004-2009 (n=406)**

| | Restricted vs. Recommended | | | | Not Recommended vs. Recommended | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.035 | 0.198 | -0.089 | 0.018 | -0.103 | 0.052 | -0.206 | 0.001 |
| RCT duration of follow-up | -0.004 | 0.196 | -0.009 | 0.002 | -0.009 | 0.011 | -0.015 | -0.002 |
| Use of active comparator in RCT | -0.708 | 0.050 | -1.415 | -0.001 | -0.897 | 0.022 | -1.662 | -0.131 |
| Clinical superiority demonstrated in RCT | -0.523 | 0.111 | -1.166 | 0.120 | -0.352 | 0.323 | -1.050 | 0.346 |
| ICER | 0.000025 | 0.011 | 0.0000057 | 0.0000448 | 0.000033 | 0.001 | 0.000013 | 0.000053 |
| Size of eligible patient population | -0.00000036 | 0.405 | -0.0000012 | 0.0000005 | -0.00000079 | 0.332 | -0.0000024 | 0.0000008 |
| Number of technologies appraised simultaneously | 0.648 | 0.001 | 0.277 | 1.018 | 0.466 | 0.074 | -0.044 | 0.977 |
| National population size | 0.00000094 | 0.087 | -0.00000014 | 0.0000020 | 0.00000174 | 0.029 | 0.00000018 | 0.0000033 |
| Central nervous system | 0.882 | 0.319 | -0.852 | 2.616 | 1.356 | 0.151 | -0.493 | 3.205 |
| Malignancy/immunosuppression therapy | 0.906 | 0.295 | -0.788 | 2.600 | 1.282 | 0.171 | -0.553 | 3.117 |
| Respiratory system | 1.415 | 0.302 | -1.271 | 4.102 | 2.172 | 0.118 | -0.554 | 4.899 |
| Cardiovascular disease | 0.355 | 0.685 | -1.359 | 2.070 | 0.453 | 0.638 | -1.431 | 2.336 |
| Endocrine system | 0.443 | 0.647 | -1.453 | 2.339 | 0.263 | 0.802 | -1.790 | 2.316 |
| Gastro-intestinal disorders | 0.240 | 0.869 | -2.615 | 3.095 | 2.001 | 0.141 | -0.661 | 4.664 |
| Infections | 0.363 | 0.678 | -1.348 | 2.074 | -0.063 | 0.949 | -1.976 | 1.851 |
| musculoskeletal and joint diseases | 0.348 | 0.721 | -1.561 | 2.256 | 0.290 | 0.786 | -1.806 | 2.386 |
| Nutrition and blood | 0.224 | 0.835 | -1.879 | 2.327 | 1.033 | 0.351 | -1.137 | 3.204 |
| Skin | 0.566 | 0.564 | -1.358 | 2.490 | -0.249 | 0.828 | -2.498 | 1.999 |
| SMC | 46.499 | 0.083 | -6.060 | 99.059 | 86.641 | 0.027 | 10.061 | 163.222 |
| Constant | -51.408 | 0.083 | -109.603 | 6.786 | -95.335 | 0.027 | -180.057 | -10.613 |

Impact of alternative model specifications: sensitivity analyses

Sensitivity analyses were conducted on the multivariate analysis of the pooled sample. This included: i) examining the impact of a binary rather than three-category outcome variable, ii) restricting the base-case analysis to observations related to cancer therapies, and iii) restricting the base-case analysis to observations related to technologies common across the four HTA bodies included in the analysis.

A sensitivity analysis was conducted using a binary outcome variable. The model using a binary outcome category provided similar results to the base-case pooled analysis (Table 8.6). The R-squared was 0.14, suggesting that this model specification could explain 14% of variability in a pooled sample of coverage decisions, similar to the pseudo R-squared obtained in the base-case model (0.13). The impact of the factors observed in the base-case model was reflected in this binary model to a large extent, although there were some variables for which the impact on coverage decisions was no longer statistically significant (e.g. disease areas including technologies for eye-related disorders, or cancer).

Clinical trial results and the disease area continued to play a role in this sensitivity analysis as they did in the base-case analysis. As observed in the base-case analysis, demonstration of superiority in the clinical trials increased the odds of a coverage relative to non-coverage ($p=0.004$). The disease area in which the technology was indicated did not play as clear a role in this sensitivity analysis as in the base case model. Of the five disease areas that were found to have a significant effect on the base-case analysis, two remained significant – technologies for respiratory diseases and for musculoskeletal/joint diseases. The former decreased the odds of coverage, while the latter increased the odds of coverage. The impact of cancer therapies on the odds of coverage was no longer observed in this binary analysis. The economic variables were not found to have a significant effect in this binary analysis.

Process and socio-economic characteristics remained important in this binary analysis as in the base-case analysis. The inclusion of patient group submissions continued to be an important factor, decreasing the odds of coverage ($p=0.014$). A unit increase in the number of technologies appraised simultaneously appeared to increase the odds of coverage relative to no coverage, but was not statistically significant ($p=0.247$). The

population size served by the HTA body continued to impact significantly in this sensitivity analysis: a unit increase in the population size decreased the odds of coverage ($p<0.0001$). Finally, the effect of the HTA body, all other factors held constant, was also observed in this binary outcome model. NICE and HAS were associated with increased odds of coverage ($p<0.001$), while the SMC was linked to significantly decreasing the odds of coverage ($p<0.001$).

In general, the binary model confirmed the results of the base-case 1 analysis in that the majority of factors found to impact on coverage decisions in the base-case analysis also had an impact in this model using binary outcome categories. However, use of binary outcome reduced access to information about whether factors behave differently when coverage is complete or restricted.

**Table 8.6 Sensitivity Analysis 1: Multivariate analysis of pooled sample of coverage decisions of NICE, SMC, CVZ, HAS 2004-2009 (base case 1): using binary outcome variable**

| | **Listed vs. Not Listed** | | | |
|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | 0.045 | 0.107 | -0.010 | 0.100 |
| RCT duration of follow-up | 0.00226 | 0.176 | -0.00102 | 0.00553 |
| Use of active comparator in RCT | 0.58 | 0.001 | 0.23 | 0.94 |
| Clinical superiority demonstrated in RCT | 0.48 | 0.004 | 0.15 | 0.82 |
| Size of eligible patient population | -0.00000025 | 0.218 | -0.00000065 | 0.00000015 |
| Orphan designation status | 0.565 | 0.038 | 0.031 | 1.098 |
| Patient submission included | -0.59 | 0.014 | -1.06 | -0.12 |
| Number of technologies appraised simultaneously | 0.232 | 0.247 | -0.161 | 0.624 |
| National population size | -0.0000007 | 0.000 | -0.0000010 | -0.0000003 |
| Central nervous system | -0.22 | 0.323 | -0.66 | 0.22 |
| Eye | 0.63 | 0.357 | -0.71 | 1.97 |
| Malignancy/immunosuppresion therapy | 0.15 | 0.494 | -0.28 | 0.58 |
| musculoskeletal and joint diseases | 0.826 | 0.015 | 0.159 | 1.493 |
| obstetrics, gynaecology, and urinary-tract disorders | 0.07 | 0.916 | -1.21 | 1.35 |
| Respiratory system | -1.45 | 0.002 | -2.36 | -0.54 |
| Cardiovascular disease | 0.30 | 0.261 | -0.22 | 0.83 |
| Skin | 0.29 | 0.451 | -0.46 | 1.04 |
| NICE | 24.73 | <0.0001 | 11.85 | 37.60 |
| SMC | -8.75 | <0.0001 | -12.59 | -4.91 |
| HAS | 29.36 | <0.0001 | 13.36 | 45.37 |
| Constant | 11.49 | <0.0001 | 5.87 | 17.11 |

In a second sensitivity analysis, a pooled analysis was also conducted on the sample of technologies indicated for cancer treatment. As highlighted in the descriptive analyses, HTA bodies varied in the nature of the diseases for which their respective technologies were indicated. It was hypothesised that this underlying variation in the disease areas

appraised could act as an important confounder in the analyses and lead to variation in the HTA coverage decisions. To address this concern, in the base-case analysis specific variables reflecting BNF disease categories were used to control for such variation. In this sensitivity analysis, an alternative method was applied that restricted the analysis of coverage decisions to a specific disease area. Cancer therapy was selected as a pertinent disease area to focus on, and technologies with BNF 8 category were considered as a proxy. Of the 977 appraisals, 247 included cancer therapies. The coverage decisions made for cancer therapies were modelled using a multinomial logit regression, where coverage decisions were regressed against the same set of explanatory variables as was used in the base-case analysis. The aim of this analysis was to understand whether those factors driving decision-making across the four HTA bodies vary by disease area, and also to increase consistency of the sample upon which the analysis was made by focusing on a specific therapy area.

The resulting multinomial logit model of coverage decisions for cancer therapies across four HTA bodies is shown in Table 8.7. The pseudo R-squared was 0.24, suggesting that this model specification could explain approximately 24% of variability in a pooled sample of coverage decisions, almost double the pseudo R-squared obtained in the base-case model (0.13). The clinical factors of relevance in the base-case model continued to be of importance in this model restricted to cancer therapies. An increase in the number and duration of RCTs, the use of an active comparator and the demonstration of superior efficacy were all found to increase the odds of recommendation relative to non-recommendation, the majority of which were statistically significant. In terms of disease characteristics, orphan designation increased the odds of a recommendation relative to non-recommendation, and those effects were statistically significant ($p=0.014$). Increasing the size of the eligible patient population increased the odds of a restriction and non-recommendation, the latter effect was statistically significant ($p=0.001$).

Process and socio-economic factors identified in the base-case model remained significant in this sensitivity analysis. Process factors present in the base-case model, namely the use of patient submissions and number of technologies appraised, had a similar impact in this model. The presence of patient submissions increased the odds of restriction and non-recommendation (statistically significant in the latter with $p=0.014$). As more technologies were appraised simultaneously this increased the odds of a

restriction and non-recommendation, although this effect was not statistically significant. Finally, as a reflection of socio-economic factors, a unit increase in the size of the population covered by the HTA body increased the odds of a restriction (p=0.255) and non-recommendation (p=0.066). The effect of the HTA body observed in this sensitivity analysis was consistent with that observed in the base case 1 model.

Thus, the results of the pooled sensitivity analyses focusing on coverage decisions for cancer therapies by the four HTA bodies suggests that the factors driving these decisions are similar to those factors explaining variation in coverage decisions in the total pooled sample. In addition, the factors within this sensitivity analysis were able to explain 24% of variability in coverage decisions compared with 13% in the base case model.

**Table 8.7 Sensitivity Analysis 2: Multivariate analysis of pooled sample of coverage decisions of NICE, SMC, CVZ, HAS 2004-2009: sensivitiy analysis using sub-sample of cancer therapies**

| | Restricted vs. Recommended | | | | Not Recommended vs. Recommended | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.155 | 0.217 | -0.402 | 0.0914 | -0.308 | 0.043 | -0.605 | -0.0102 |
| RCT duration of follow-up | -0.002 | 0.462 | -0.0071 | 0.00321 | -0.009 | 0.019 | -0.0157 | -0.00139 |
| Use of active comparator in RCT | -0.491 | 0.234 | -1.298 | 0.316 | -0.717 | 0.110 | -1.597 | 0.162 |
| Clinical superiority demonstrated in RCT | 0.556 | 0.187 | -0.269 | 1.381 | -0.641 | 0.160 | -1.534 | 0.252 |
| Size of eligible patient population | 0.00000444 | 0.589 | -0.000011700 | 0.00002050 | 0.00002290 | 0.001 | 0.000008900 | 0.00003680 |
| Orphan designation status | -0.714 | 0.142 | -1.668 | 0.240 | -1.406 | 0.014 | -2.532 | -0.280 |
| Patient submission included | -0.026 | 0.969 | -1.338 | 1.287 | 1.686 | 0.014 | 0.343 | 3.030 |
| Number of technologies appraised simultaneously | 0.431 | 0.200 | -0.228 | 1.090 | -0.100 | 0.840 | -1.070 | 0.870 |
| National population size | 0.0000005 | 0.255 | -0.0000003 | 0.0000013 | 0.0000008 | 0.066 | -0.0000001 | 0.0000017 |
| NICE | -15.524 | 0.321 | -46.178 | 15.130 | -29.899 | 0.081 | -63.486 | 3.688 |
| SMC | 8.172 | 0.080 | -0.980 | 17.323 | 12.657 | 0.014 | 2.574 | 22.739 |
| HAS | -19.698 | 0.309 | -57.663 | 18.268 | -36.070 | 0.090 | -77.746 | 5.606 |
| Constant | -9.687 | 0.158 | -23.127 | 3.752 | -14.584 | 0.054 | -29.439 | 0.272 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.25.

In the third sensitivity-analysis, a sub-set of the pooled sample was utilised which included those technologies that had been appraised by all four HTA bodies at least once. This sub-sample included a total of 192 appraisals of 26 technologies from 2004-2009. The rationale for this sub-analysis was to create a more homogeneous platform for comparison across HTA bodies, by selecting those technologies in common among them. The hypothesis was that some of the variation in coverage observed between HTA bodies could be due to differences in the technologies appraised. By focusing the analysis on those technologies reviewed across all four HTA bodies, the nature of the factors driving decisions within this sample could be assessed and compared with the base-case analysis.

The regression model of the sub-sample of technologies appraised by all four HTA states was associated with a pseudo R-squared which was more than double that of the base-case analysis (0.27 in this sub-analysis and 0.13 in the base-case pooled analysis) (Table 8.8). As in the base-case pooled analysis, the demonstration of superiority in the clinical trial primary endpoint in this sub-sample statistically significantly decreased the log-odds of a restriction and non-recommendation relative to a recommendation. A unit increase in the number of RCTs and duration of the RCT appraised decreased the log odds of a non-recommendation, and this was statistically significant (p=0.002, p=0.012). The effect of the number of RCTs and duration of the RCT on the odds of a restriction was not statistically significant. The use of an active comparator within the clinical trial(s) decreased the odds of a restriction (p=0.103) and for non-recommendation (p=0.161).

**Table 8.8 Sensivity Analysis 3: Multivariate analysis of pooled sample of coverage decisions of NICE, SMC, CVZ, HAS 2004-2009: sensivitiy analysis using sub-sample of common technologies**

| | Restricted vs. Recommended | | | | Not Recommended vs. Recommended | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.200 | 0.131 | -0.460 | 0.060 | -0.612 | 0.002 | -1.009 | -0.216 |
| RCT duration of follow-up | -0.002 | 0.772 | -0.012 | 0.0088 | -0.020 | 0.012 | -0.035 | -0.0044 |
| Use of active comparator in RCT | -0.802 | 0.103 | -1.764 | 0.161 | -0.814 | 0.161 | -1.951 | 0.323 |
| Clinical superiority demonstrated in RCT | -1.005 | 0.054 | -2.025 | 0.016 | -1.383 | 0.032 | -2.645 | -0.122 |
| Size of eligible patient population | 0.0000043 | 0.273 | -0.0000034 | 0.000012 | 0.000012 | 0.019 | 0.0000019 | 0.000021 |
| Orphan designation status | 0.484 | 0.544 | -1.078 | 2.047 | -0.410 | 0.683 | -2.377 | 1.557 |
| Patient submission included | 0.448 | 0.584 | -1.156 | 2.052 | 1.698 | 0.055 | -0.038 | 3.435 |
| Number of technologies appraised simultaneously | 0.858 | 0.098 | -0.159 | 1.875 | 1.153 | 0.054 | -0.019 | 2.325 |
| National population size | 0.000 | 0.006 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 |
| Central nervous system | 0.433 | 0.685 | -1.662 | 2.528 | -2.044 | 0.081 | -4.340 | 0.253 |
| Malignancy/immunosuppresion therapy | 0.624 | 0.395 | -0.815 | 2.064 | -0.868 | 0.261 | -2.382 | 0.646 |
| musculoskeletal and joint diseases | 0.481 | 0.550 | -1.098 | 2.060 | -2.821 | 0.007 | -4.864 | -0.778 |
| Cardiovascular disease | 0.495 | 0.576 | -1.240 | 2.230 | -4.067 | 0.007 | -7.006 | -1.129 |
| Skin | 1.531 | 0.161 | -0.609 | 3.672 | -1.424 | 0.291 | -4.069 | 1.222 |
| NICE | -56.052 | 0.007 | -96.555 | -15.549 | -86.996 | 0.003 | -143.759 | -30.233 |
| SMC | 18.646 | 0.002 | 6.563 | 30.728 | 30.322 | 0.000 | 13.314 | 47.331 |
| HAS | -70.036 | 0.006 | -120.233 | -19.839 | -108.219 | 0.003 | -178.755 | -37.683 |
| Constant | -25.532 | 0.005 | -43.260 | -7.804 | -38.376 | 0.002 | -63.052 | -13.700 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.30.

Aside from clinical and economic factors, two process factors, namely the use of patient submissions and the number of technologies appraised were also significant in explaining variation in coverage decisions in this sub-sample. The inclusion of patient group submissions within the appraisal process increased the log-odds of a non recommendation versus a recommendation (p=0.055). The number of technologies appraised statistically significantly increased the log-odds of restriction or non-recommendation(p=0.098, p=0.054, respectively).

In terms of socio-economic factors, the national population which falls within the HTA's remit appeared to impact on coverage decisions: a unit increase in the population size increased the odds of a restriction (p=0.006) and non-recommendation (p=0.002). Finally, when examining the role of the HTA body, all other factors held constant, NICE and HAS increased the log-odds of recommendation, while the SMC increased the odds of restriction and non-recommendation. These observations were statistically significant and consistent with the effects observed in base case model 1.

Overall, the results of this third sensitivity analysis suggest that compared to the base case model 1, the explanatory power of the combination of clinical, process and socio-economic factors is higher in the sensitivity analysis including only those technologies that all four HTA bodies hold in common (as implied by the higher pseudo R-squared).

## 8.4 Discussion

The aim of this chapter was to model the factors driving coverage decisions observed within the pooled data set of appraisals performed by four European HTA agencies: NICE, SMC, CVZ and CVZ. Analysis of pooled coverage decisions across HTA bodies is characterised by significant challenges. However, there were two advantages to running a pooled analysis. The first benefit is the maximisation of sample size to examine the role of factors in decision-making. The combined data set consisted of 977 appraisals, representing currently the largest dataset of HTA appraisals in Europe, compared with analyses available in the published literature. Secondly, the data set was created by extracting information on the selection of explanatory variables from each appraisal and each HTA body by the same researcher. This, coupled with the use of a formal extraction protocol described in Chapter 3, increased consistency in how the data were extracted.

### *8.4.1 Pattern of coverage decisions in pooled analysis*

The evolution in coverage decisions by NICE, SMC, CVZ and HAS was examined through descriptive analyses which mapped the yearly patterns of coverage decisions during the period 2004-2009. This showed firstly that the pattern of coverage decision-making evolved over time, generally moving towards increasing levels of restriction and non-recommendation primarily within the NICE and HAS samples. In addition, the descriptive analyses highlighted similarities and differences between HTA bodies in the characteristics of the recommended, restricted and non recommended technologies.

### *8.4.2 The impact of clinical variables and disease characteristics on decision-making within a pooled HTA analysis*

In general, clinical and disease explanatory variables, reflecting higher quality clinical evidence and specific disease characteristics, tended to increase the log-odds of a recommendation versus a restriction or non recommendation. Orphan designation statistically significantly increased the log odds of recommendation relative to non-recommendation, and technologies indicated for cancer or eye disorders were associated with statistically significantly increased log-odds of recommendation relative to a restriction and non-recommendation in base case 1 model. No statistically significant effect of orphan designation was noted in base case model 3 which assessed pooled decisions from NICE and SMC. The results of the pooled analysis were consistent with the observed effects of clinical and disease characteristics within the individual HTA analyses.

### *8.4.3 The impact of economic evidence on decision-making within a pooled HTA analysis*

A specific base case analysis was presented to evaluate the role of economic variables, in particular the ICER, amongst the HTA bodies that routinely consider this information in their appraisal process. The pooled analysis of NICE and SMC appraisals confirmed the hypothesis that the effect of the ICER is similar to that observed in the individual HTA analysis. That is, an increasing ICER increases the log odds of restriction and non-recommendation relative to recommendation, and this effect is highly statistically significant. This analysis represents one of the largest analyses of the effect of cost-effectiveness on HTA decision-making with a sample of over 400 appraisals.

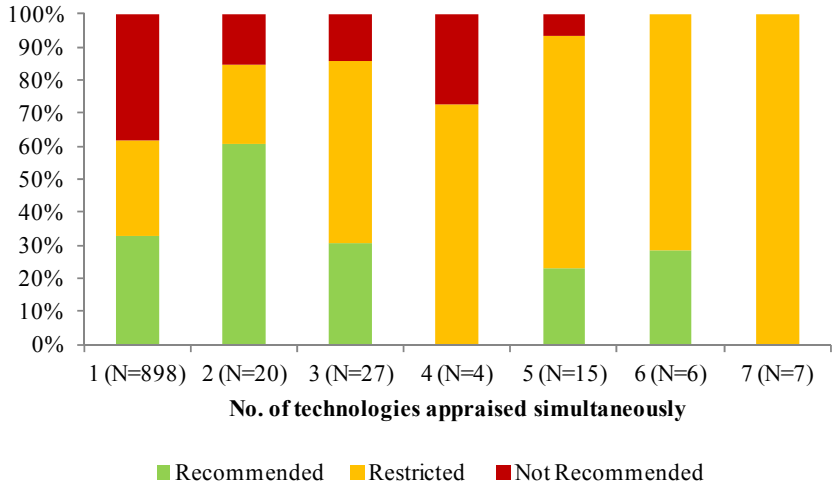
### *8.4.4 The impact of the appraisal process on decision-making within a pooled HTA analysis*

Evidence of the impact of patient group submissions on coverage decision-making within this pooled analysis supports the hypothesis of the increasing role of process factors in explaining decision-making within a pooled analysis. HTA bodies varied in the use of evidence provided from patient group submissions. While patient group submissions were commonly included as part of the evidence appraised by NICE (in 87% of appraisals) but less so by the other HTA bodies (42% in SMC decision-making, 0% HAS submissions and 4% within CVZ appraisals). These differences in use of evidence from patient group submissions was statistically significant ($p<0.0001$). Within the multivariate analysis, the availability of patient submissions increased the odds of restriction relative to recommendation in the base case (1) analysis. Appraisals supported by evidence from patient submissions tended to be for technologies indicated for cancer treatment (32% of appraisals vs. 23% for appraisals not supported by cancer therapies), and representing on average a smaller target patient population (78,858 patients vs. 106,289 patients, on average). Whether supported or not supported by patient evidence, the characteristics of the RCTs, the year of appraisal and the proportion of orphan designations, and the availability of alternative therapies were similar.

An additional process aspect relates to the number of technologies that can be appraised. NICE regularly allows the appraisal of multiple technologies simultaneously, while for other HTA bodies this is rare (CVZ, HAS) or not part of the process (SMC). The multivariate analyses confirmed that an increase in the number of technologies appraised simultaneously statistically significantly increased the odds of restriction relative to recommendation in both base case models and a statistically significant effect on the odds of non-recommendation relative to recommendation in the base case model 3 which includes NICE and SMC appraisals. Descriptive analysis of the distribution of coverage decisions according to the number of technologies appraised simultaneously illustrates the trend towards increasing restriction as the number of technologies appraised increases (Figure 8.6, note small sample sizes in certain categories) Barbieri et al. (2009), in their comparison of NICE and SMC coverage decisions, attempted to explain differences in coverage decisions through the fact that NICE uses third party assessment, while SMC does not. However, the lack of an adequate sample size impeded the authors from firstly demonstrating the presence of statistically significant

differences in coverage decisions between the two bodies, and secondly, from demonstrating if these differences were driven by third party technology assessment processes, rather than other factors (e.g. use of clinical/economic evidence, the agency's mission, other process elements such as inclusion of patient groups etc).

**Figure 8.6 NICE, SMC, CVZ and HAS coverage decisions stratified by number of technologies appraised simultaneously (n=977)**




### *8.4.5 The impact of socio-economic context on decision-making within a pooled HTA analysis*

National population size was shown to impact significantly on HTA decision-making in this pooled analysis, an effect which was not previously observed in the individual analysis and which supports the hypothesis of the increasing role of socio-economic factors in explaining decision-making within a pooled analysis. An increasing population size increased the odds of both restriction and non-recommendation relative to recommendation, and the effects were statistically significant in both base case models. It is not clear to what extent this result reflects the effect of the absolute size of the patient population eligible for treatment – this variable also increases the odds of restriction and non-recommendation but its effect is only statistically significant in base case model 1 (on odds of restriction relative to recommendation).

### *8.4.6 The "HTA body effect"*

The results of the base case 1 analyses provided evidence that confirm the hypothesis of a strong "HTA body effect" impacting on decision-outcomes, even when adjusting for a range of confounding factors. The impact of the HTA body was highly statistically significant across all assessments. Relative to the CVZ, the results suggest that NICE and HAS assessment bodies are associated with a decreased odds of a restriction or non

recommendation relative to recommendation, while the SMC is associated with an increased log odds of both restriction and non-recommendation relative to recommendation.

While not examining across multiple *European* HTA agencies, Clement et al. (2009) produced a thorough descriptive analysis comparing NICE with Australian and Canadian HTA agencies, obtaining results that are relevant to this research. The results of Clement et al. (2009) suggest that there are statistically significant differences in the nature of the coverage decisions made by NICE, PBAC and CDR. Statistically significant differences in the nature of coverage decisions were also observed between the four HTA agencies considered in this thesis. Additionally, the results show that there are differences in the characteristics of the technologies appraised by the agencies, and the evidence that supports them (Clement et al. 2009). This was also a finding in the pooled descriptive analyses of this research. However, the study by Clement et al. (2009) was not able to clarify the relative importance of each of the examined factors, as a regression model was not attempted to allow for adjustment of the effect of individual factors on one another. Nor was the study able to provide direct evidence of the impact of variation in a specific variable leading to variation in coverage decisions. The multivariate analysis conducted in this thesis therefore provides additional evidence that examines the role of individual factors on coverage decisions made across HTA bodies, while adjusting for the presence of other characteristics/factors known to impact on decision-making.

### *8.4.7 Limitations*

Significant limitations must be taken into account when interpreting the results of this pooled analysis. A significant challenge in conducting pooled analyses of coverage decisions is that across the four HTA bodies analysed, coverage decision are not communicated in the same way. For instance HAS uses ASMR ratings to define incremental therapeutic value, CVZ attributes technologies to different reimbursement lists, SMC uses a three-category coverage system and NICE a binary category system. Aside from differences in the communication of coverage decisions, the concept of a 'restricted' category is heterogeneous both within and across HTA bodies and thus suggests that there may be a limitation in using a three category outcome variable that uses 'restricted' technologies as one such category. To attempt to address this limitation, a sensitivity analysis was implemented using a binary outcome variable that

combined recommended and restricted categories into a single 'covered' category, and not recommended decisions as 'not covered'. The aim of this sensitivity analysis was to evaluate to what degree changes in the classification of the outcome variable altered the role of factors from the base-case model. The model using a binary outcome category provided very similar results to the base-case pooled analysis. In addition, the sensitivity analysis suggested that use of a binary outcome variable reduced access to useful information about how the role of factors may change between different 'degrees' of coverage: i.e. between recommended and restricted technologies. Thus, based on the available analyses conducted, it would seem that although there is heterogeneity between HTA bodies in how coverage decisions are communicated and what they consist of (e.g. restricted decisions), the use of a three-category outcome variable provides a greater level of insight into the 'behaviour' of factors between different types of coverage decisions, compared to use of a binary outcome variable.

When combining coverage decisions across HTA agencies, it is challenging to create as homogenous a platform as possible upon which to model. A key limitation identified in this research is that HTA bodies utilise different criteria to define the sample of technologies they appraise. The SMC, HAS and CVZ review all new technologies and new indications of existing technologies. CVZ also reviews existing technologies for non-licensed indications. NICE technologies included for appraisal are the result of a complex selection process involving the Department of Health and other stakeholders, and is not driven solely by the availability of a new technology or a new indication. In addition, NICE, HAS and CVZ re-review a sub-set of the technologies after a specified time period, which is not done by the SMC. This resulted in a situation in which, in the pooled sample, less than 10% of 348 technologies were reviewed by all four HTA bodies. To overcome the potential limitations associated with this, two approaches were tested. In the base case model, the heterogeneity of the samples across HTA bodies was adjusted for by including in the model variables that captured the nature of the disease corresponding to the technologies appraised. In addition, two sensitivity analyses were conducted that restricted the sample upon which the multivariate analyses were performed. A specific analysis was conducted on the sub-population of technologies reviewed by all four agencies, representing a total of 192 appraisals. The aim of this sub-analysis was to create a more homogenous platform to better estimate the impact of the selected variables on the likelihood of a recommendation versus a restriction or non recommendation. Overall, the results of this sub-analysis show that

the explanatory power of the combination of clinical, economic, process and socio-economic factors is higher in the sub-analysis including only those technologies appraised by all four HTA bodies. This may suggest that increasing the homogeneity of the sample facilitates a more robust analysis of the role of explanatory variables on coverage decisions. On the other hand, homogeneity is obtained at the expense of significantly restricting the sample of analysis to less than 20% of the total available sample, increasing the odds of bias in the sample.

Another means of creating a more homogeneous platform for comparison was tested by limiting the analysis to appraisals performed within a specific disease area: in this case, cancer. This analysis was performed as an alternative means of addressing the limitation that variability in the technologies appraised due to differing selection process between HTA states could reduce the usefulness and generalisability of the results. To achieve this aim, 247 appraisals were identified that appraised technologies for the treatment of cancer. The results of this analysis suggest that the factors driving these decisions may be different from those that have been found to be of significance in the total pooled sample. In particular, in the cancer therapy sub-population, there appears to be a greater impact of process-related characteristics, rather than clinical characteristics. Thus, it may be appropriate to consider that HTA bodies may not utilise the same criteria in their decision-making across all disease areas.

Overall, the results of the quantitative analysis of pooled coverage decisions made by NICE, SMC, CVZ and HAS was useful to demonstrate that a combination of evidence, process and socio-economic context factors help explain variability in coverage decisions made across these four HTA bodies. The analyses also represent the first model of coverage decisions across four European HTA bodies based on 30+ explanatory variables extracted directly from their appraisals. This analysis suggests that those factors of importance in explaining variation in coverage decisions across HTA bodies may not be the same as those factors explaining variation in coverage decisions within each HTA body – namely, process and socio-economic context factors play a more evident role in this analysis than in the single HTA analyses. At the same time, the impact of the clinical and disease factors observed in the single analysis remains very significant in the pooled analysis. Across all HTA bodies, recommended technologies tended to have higher quality clinical data (higher number of clinical trials of larger size and longer duration, with use of an active comparator). The descriptive

analyses suggested that there was most variation between HTA bodies in the restricted technologies, and least variation between HTA bodies in the technologies that were not recommended.

The pooled analyses provide new evidence that differences exist between HTA bodies in the nature of technologies appraised and the clinical evidence considered, the process through which HTA bodies appraised the technologies and the socio-economic context in which the appraisals took place. Importantly, the evidence also shows that the HTA body effect is present which contributes significantly to decision-making, even when adjusting for a range of confounding variables. The analyses also emphasise that, amongst those HTA bodies that consider cost-effectiveness analyses, the ICER has a significant effect on the coverage decision. The pooled analysis has also provided an opportunity to examine the role of process and socio-economic factors on decision-making. Having examined the factors driving coverage decisions both at individual HTA body level and across agencies, Chapter 9 concludes by reflecting on the results presented and their policy implications.

## 8.5 References

Kanavos, Panos, Elena Nicod, Stacey van den Aardweg, and Stephen Pomedli. 2010. The impact of health technology assessments: an international comparison. *Euro Observer* (4).

Raftery, J. 2006. Review of NICE's recommendations, 1999-2005. *BMJ* 332 (7552):1266-8.

# 9 Conclusion

> "Each [EU Member State] has developed its own approach to maintaining a regulatory equilibrium between public health, healthcare and industrial policy interests, reflecting particular national circumstances and requirements."
> (Permanand 2006 p. 5)

> "…patients across Europe do not have equitable access to new innovative cancer drugs." (Wilking and Jonsson 2005 p. 91)

## 9.1 Thesis Overview

Within European healthcare systems, coverage decisions represent a key decision point within the complex process that governs funding and access for pharmaceuticals. The interest in analysing coverage decisions made by European HTA bodies is related to the realisation that, ultimately, such decisions not only impact on the technology's price and volume, but also impact on patient access to treatment. As demand for healthcare increases while governments and healthcare providers struggle to manage healthcare expenditure, the need for difficult decision-making with regard to healthcare spending and allocation of resources is amplified, and tensions arise between HTA agencies, patients, providers, healthcare policy makers and manufacturers as coverage decisions are made. In this context of difficult decision-making, there is a recognized need for a greater understanding of the processes and criteria adopted by HTA bodies in making decisions about the public funding of pharmaceutical technologies.

HTA decision-making poses an interesting paradox. On the one hand, it is characterised by the desire to rely on rigorous scientific criteria to justify its coverage decision (e.g. NICE 2008); on the other hand, it is characterised by the need to serve the healthcare system in which it operates, governed by public health, health policy and industrial policy concerns relative to the national context (Permanand 2006). Hand-in-hand with this apparent paradox, existing theoretical models propose alternative perspectives on the drivers of HTA coverage decisions (Weiss 1979): namely the 'problem-solving model' and the 'interactive model'. The 'problem-solving model' represents a world in which decision-making is driven by bespoke evidence resulting from aligned interaction between decision-maker and researcher. The latter model, the 'interactive model', proposes that decision-making is driven by interactions between different stakeholders

and factors within the decision-making process, of which evidence is only one component. The question that follows from this is: where do European HTA bodies sit within these theoretical models of decision-making? Are they primarily driven by evidence, process or context, or influenced by a combination of factors?

This thesis has aimed to analyse the factors that drive HTA coverage decisions on pharmaceuticals in a selection of EU Member States, focusing specifically on NICE and SMC in the United Kingdom, alongside the CVZ in the Netherlands and HAS in France. The thesis has aimed to address this research question by developing an appropriate evidence base and a corresponding set of methods that created a robust platform for analysing the extent to which evidence, process and context factors influence coverage decisions. An important stepping stone in identifying the focus and methods for this research was the analysis of the literature to gain an awareness of the current knowledge and understanding of factors driving coverage decisions by HTA bodies in Europe.

Following an introductory chapter, Chapter 2 identified the gaps in the literature that this thesis sought to address, and provided the analytical framework for the research. The chapter on methods (Chapter 3) described the process through which a bespoke dataset was created, and the statistical methods adopted to analyse the factors that had been identified as potential determinants of coverage decisions. Chapters 4 to 7 provided analyses of coverage decisions for each individual HTA body, while Chapter 8 set out an analysis of pooled coverage decisions from all four HTA bodies. This chapter discusses the overall results in more detail and outlines some policy implications and areas that would warrant further research.

**Table 9.1 Summary of statistically significant explanatory variables in descriptive analyses (p≤ 0.05) and multivariate analyses (p≤ 0.10) in NICE, SMC, CVZ and HAS coverage decisions made between 2004-2009**

| Explanatory Variables | NICE | SMC | CVZ | HAS |
|---|---|---|---|---|
| **Component 1 - Evidence** *Clinical Package* | No. of RCTs<br>Size of RCT population<br>Duration of RCT<br>**Statistical Significance**<br>**Superiority demonstrated in RCT**<br>Active comparator used<br>Consideration of observational data | **Size of RCT population**<br>**Duration of RCT**<br>Active comparator used<br>Superiority not demonstrated | Number of RCT<br>Size of RCT<br>**RCT duration (lack of data)**<br>**Superiority demonstrated**<br>**Use of active comparator** | No of RCTs<br>**RCT demonstrates superiority**<br>Duration of RCT<br>Active comparator used |
| *Economic Package* | **ICER**<br>Range of ICERs<br>Uncertainty around ICER: probabilistic, univariate<br>Use of non-CUA economic models<br>Maximum budget impact | **ICER**<br>Uncertainty around ICER: probabilistic, univariate | Non-CUA analyses submitted<br>**Budget Impact** | -- |
| *Disease characteristics* | Prevalence of disease<br>Technologies for treatment of CNS disorders | **Prevalence of disease**<br>**Alternative**<br>Orphan designated status<br>**BNF categories (infectious diseases, skin diseases)** | **Prevalence of disease**<br>**BNF categories (therapies for malignancies, cardiovascular diseases, obstetrics/gynaecology)** | **Prevalence of disease**<br>**Availability of alternatives**<br>**Orphan status**<br>**BNF category** |
| **Component 2 - Process** | Use of STA vs. MTA process<br>**No. of technologies reviewed simultaneously** | - | **Presence of patient submission**<br>Priority disease<br>Request for future CEA<br>Patient co-payment | Level of reimbursement<br>Hospital use only<br>Request for post-launch study<br>Reason for request |
| **Component 3 - Context** | **Date of appraisal**<br>**National population size**<br>Percentage GDP spent on healthcare<br>Pharmaceutical expenditure per patient per year | Date of Review | Date of Review<br>Pharmaceutical expenditure per patient per year<br>**National population size** | Date of Review<br>**Pharmaceutical Expenditure per patient per year**<br>**National population size** Election year |

Note: Variables highlighted in bold text had a statistically significant impact on coverage decisions in the multivariate analyses.

## 9.2 Discussion of Results

### *9.2.1 NICE*

> "Our job is to make sure taxpayers' money is only spent on healthcare that works and is good value" (NICE statement in BBC Breakfast News, 13 May 2008)
>
> "It's a system of blocking. They're [NICE] not looking at patients and saying "how can we fund it?", they are saying, "how can we not fund it?"…[Cancer] medicines are licensed and working yet we in Britain aren't allowed to access them. While we wait for Nice to decide, patients are dying." (Kate Spall[1] on BBC Breakfast News 13th May 2008)

The above excerpts provide two contrasting perspectives on the role of NICE and its impact on funding of technologies within the NHS in England and Wales. The assessment of NICE's decision-making in 2004-2009 presented here, and performed through descriptive and multivariate analyses, provides useful glimpses into the factors driving its decision-making. The analysis of NICE coverage decisions involved the review of appraisals of 118 technologies performed in 2004-2009. The majority of NICE coverage decisions involved restricting funding[2] to the appraised technology (58% of technologies were restricted), while the least common coverage decision was non-recommendation (14% of technologies were not recommended by NICE for NHS funding). Therefore, the pattern of coverage decisions issued by NICE, in which either full or restricted coverage is advocated more than 80% of the time, does not substantiate the claim made in the second quote cited above, that NICE is refusing patients access to new technologies in a generalised way. However, it should be noted that, when the pattern of NICE coverage decisions is examined across time, there has been a trend towards an increasing proportion of restrictions and non-recommendations relative to recommendations (see Chapters 4 and 8).

This pattern is not dissimilar to that reported in Kanavos et al. (2010) which examined NICE coverage decisions in 2007-2009: that analysis revealed that of the technologies appraised, 19% were recommended, 63% were restricted and 18% were not recommended. Clement et al. (2009) examined NICE coverage decisions between

---

[1] Kate Spall is a member of the public who has been involved in helping more than 50 patients across England and Wales to receive funding from local Primary Care Trusts ( PCTs) for new cancer drugs that had not yet been reviewed by NICE.

[2] Technologies that were restricted by NICE ranged from major restrictions where the use of the technology was limited within a sub-set of the licensed population, to minor restrictions such as the need to monitor the use of the technology.

2001- 2008 and reported 87% of technologies as listed (recommended or restricted), leaving 13% of technologies as not recommended. In the analysis of NICE decision-making during 2000-2003 produced by Dakin et al. (2006) it was reported that 21% of technologies were recommended for routine use, 66% for restricted use and 13% were not recommended. Devlin and Parkin (2004) reported NICE outcomes using a binary outcome variable, and during the period 2000-2002 reported that 71% of appraisals recommended use of the technology and 29% of appraisals did not recommend use of the technology. Therefore, the coverage pattern observed within the data set used in this thesis seems similar to that observed in other reports, suggesting that the method of classification and data extraction used was robust.

As highlighted by the results presented in Chapter 4, the characteristics of recommended, restricted and non-recommended technologies were compared descriptively, and they were found to differ significantly in terms of clinical and economic characteristics, as well as in relation to the process and context in which the decisions were made. Guided by the results of the descriptive analyses, multivariate analyses were performed to examine the relative role of various explanatory variables in explaining NICE decision-making. When NICE coverage decisions were analysed through multivariate analyses, the results suggested that variation in NICE coverage decisions could be explained by four variables: whether statistical superiority of the primary endpoint in the RCT was demonstrated, the incremental cost-effectiveness ratio, the number of technologies considered within the appraisal, and the year of appraisal (Table 9.1).

In the analysis of NICE decision-making presented here, demonstration by the technology under appraisal of statistically significant superiority in its primary endpoint increased the odds of recommendation. This result can be seen to reflect the role of evidence-based medicine in coverage decisions and the fact that NICE defines the value of the compound in terms of the ability of the technology to demonstrate, with greater certainty, its incremental clinical value through superiority designed trials that provide stronger data to support a funding decision than technologies not able to provide evidence of superior efficacy. Contrasting this result with those previously published in the literature (Dakin et al. 2006; Devlin and Parkin 2004), it is noteworthy that Devlin and Parkin (2004) did not consider the demonstration of clinical superiority as a variable in their analysis while in Dakin et al. (2006), this variable was measured but

was not found to have a statistically significant impact, although the sample size was smaller than that used within the analyses presented here, and pertaining to NICE appraisals made in 2000-2003.

Other clinical variables which were tested included information on the characteristics of RCTs used in the appraisal (number, size, duration). Within descriptive analyses, these variables were found to differ between coverage decisions, at statistically significant levels. However, they were not key drivers within the multivariate analyses. This is not to say that the characteristics of RCTs are not important or not considered within the NICE appraisal process, but that within the multivariate analysis the demonstration of superiority and the contribution of other factors, such as the ICER, had a more significant impact. This result is not dissimilar to that observed in the analysis of NICE decisions by Dakin et al. (2006), in which RCT size was not found to have statistically significant effects on the odds of a restriction or non recommendation relative to a recommendation within their multivariate analysis. The number of RCTs, however, did have a statistically significant impact in the Dakin et al. (2006) model, which was not observed in the analyses presented here. This could be due to differences in the samples analysed: Dakin's analysis was smaller (n=60 vs. n=118 in this analysis) and the analysis included NICE appraisals published in a different time period (2000-2003). The analyses by Dakin et al. (2006) also show that the use of systematic literature reviews within the NICE appraisal process appear to decrease the odds of non recommendation and restriction relative to a recommendation at statistically significant levels. This variable was not examined in this analysis because it was noted in a review of appraisals that the majority of appraisals post 2004 included systematic reviews of the clinical and economic literature as part of the process.

Looking at the other major variables impacting on NICE's decision-making, the importance of the incremental cost-effectiveness ratio in this analysis highlights NICE's focus on identifying "healthcare that works and is good value". An increase in the ICER decreased the odds of a recommendation, and this was highly statistically significant. This would suggest that, in addition to the strength of the clinical data, the incremental costs and benefits associated with the technology, and the resulting ICER, play a significant role in coverage decisions by NICE, and represent the agency's focus on maximising the efficiency of public healthcare spending on technologies. The impact of the ICER observed in this analysis of NICE coverage decisions appears to

confirm the findings in the literature which have also previously found the ICER to be a key driver of NICE decision-making. Devlin and Parkin (2004) concluded from their analyses that the ICER has a significant impact on NICE decision-making. However, their use of a binary outcome variable precluded an assessment of whether the role of the ICER was important across coverage decisions. Dakin et al. (2006) used a three-category outcome variable and showed that the ICER had a significant effect on the odds of both a restriction relative to recommendation and non-recommendation relative to recommendation – which was also observed in the analyses presented here, confirming that the ICER is a highly important factor that drives NICE decision-making.

But the model developed here also suggests that there are process and context factors, beyond evidence considerations which explain NICE coverage decisions. When considering appraisal process-related factors, the results of the model suggest that an increase in the number of technologies reviewed simultaneously within the same appraisal increased the odds of a restriction relative to a recommendation. It was hypothesised that this may reflect the fact that NICE assessment processes differ according to the number of technologies under appraisal. Single technologies are evaluated using the STA process (in use since 2006) while multiple technologies are evaluated using the MTA process. The latter involves the use of third-party assessments to provide bespoke research to support NICE assessment, while the STA process relies on manufacturer submissions similar to the process undertaken by the SMC in Scotland. However, when the model specification was altered to include a variable capturing the use of MTA or STA processes, the effect was not statistically significant. Thus, the effect on coverage decisions arising from the appraisal of multiple technologies simultaneously could not be explained by the use of MTA or STA alone. It was suggested by Professor P. Littlejohns[3], that the increased odds of restriction associated with higher number of technologies appraised simultaneously may reflect an approach in which a 'winner' is picked among the technologies, with the remainder recommended for restricted use or non-recommendation. None of the published multivariate analyses of NICE decision-making examined the role of process factors, and thus a comparison with previous analyses is not possible.

---

[3] Littlejohns, Peter. Professor, Clinical and Public Health Director National Institute for Health and Clinical Excellence. Interviewee: Karin Cerri. Interviewed by telephone, on February 2nd 2011. Meeting minutes are provided in Appendix E.

In terms of the socio-economic context of NICE decision-making, the results of the model show that the year of appraisal impacted significantly on the coverage decision – moving from 2004 to 2009 increased the odds of a restriction or non-recommendation. The year of appraisal may reflect multiple socio-economic factors, including the political climate, a change in key staff of the HTA body, a change in societal preferences or the overall economic context. By way of comparison, Dakin et al. (2006) included this variable in their analyses, and also found a statistically significant effect of the time of appraisal on outcome. In particular, more recent appraisals had higher odds of non recommendation. However, Dakin et al. (2006) did not observe the effect of appraisal date on the odds of a restriction relative to a recommendation.

Having presented the results of the model it should be borne in mind that the analysis of NICE decision-making has been limited by several factors including: i) dependence on publicly available information for the creation of the dataset; ii) heterogeneity in the definition of 'restricted' technologies; and iii) challenges with data extraction.

The database constructed for these analyses of NICE decision-making was dependent on publicly available information. Thus, it is possible that considerations or rationales discussed by the appraisal committee in oral format were not captured in the documentation of the appraisal. In addition, the dependence on public information meant that in situations where the information was incomplete, it was not possible to ascertain if this was because the information was never considered in the appraisal or if it was considered but not recorded in the documentation. The presence of non-reporting reflects a lack of transparency associated with the documentation of the appraisal process. This was noted particularly with regard to the non-reporting of uncertainty information around the cost-utility/effectiveness results, both in terms of probabilistic sensitivity analyses as well as univariate sensitivity analyses. Both forms of sensitivity analysis are an integral part of NICE methods, requested as part of its reference case, and the value of parameter and decision-uncertainty information is very high (Claxton et al. 2005). Incomplete observations were imputed in the multivariate analyses and dummy variables were created that captured the instances in which a particular variable was incomplete, so as to be able to examine whether the presence or absence of

information on that variable had any explanatory value. It should be noted that none of these dummy variables had a statistically significant effect in the base case model.

There was also heterogeneity in the means through which technologies were restricted within NICE coverage decisions. The notion of restriction within NICE coverage decisions ranged from major restrictions, including restriction for use within a sub-set of the licensed indication, to minor restrictions such as the need to monitor the use of the technology. The notion of major and minor restrictions was suggested in research by Raftery (2006) on NICE guidance in which various sub-types of restrictions were presented. It is a limitation for this analysis to have such heterogeneity in the degree of restriction and hence of access to pharmaceuticals within a single category. However, the use of a third coverage category (ie. restriction) within the analysis was felt to better reflect the real-life decision-making in which there are multiple options available to HTA bodies, rather than the binary alternatives of simply listing (recommendation) or not listing (not recommending).

Challenges were also faced with regard to the creation of the dataset of NICE decision-making due to the dispersal of information between various documents, and potential for inconsistency between NICE guidance drafts, manufacturer reports and Assessment Group reports, leading to a reduction in the transparency of the information that was considered by NICE. This challenge was managed by creating a specific data extraction protocol with specific rules on the nature of the data that should be extracted and how to select the data of most relevance. For example, it was not uncommon for multiple models to be submitted and considered. In these instances, the model that provided the ICER which drove the decision-making was selected and included in the dataset.

### *9.2.2 SMC*

> "The purpose of the Scottish Medicines Consortium (SMC) is to accept for use those newly licensed drugs that clearly represent good value for money to NHS Scotland." (SMC 2011 p1)

SMC decision-making was analysed to understand the factors used to identify technologies that demonstrate "good value for money to NHS Scotland" relative to those technologies not able to demonstrate this value. The analysis of SMC decisions involved the review of 288 technology appraisals during 2005-2009. The most common coverage decision by the SMC was to not recommend the new technology for use (46%

of appraisals), followed by accept for restricted use (35% of appraisals), and accept for use (19% of appraisals). It would seem, therefore, that the SMC identified in approximately half its appraisals technologies that represented value for money. Other analyses of coverage decisions made by the SMC provide different proportions of coverage types, although this appears to be due to differences in time horizon. For example, the SMC Annual Report (SMC 2008) summarises coverage decisions for 2008: 31% accepted for use, 36% of technologies accepted for restricted use, and 33% technologies not recommended for use. In a similar exercise but looking at decisions in the period 2007-2009, Kanavos et al. (2010) find that the SMC recommended 28% of technologies, restricted 40% of them and did not recommend 32% of technologies. Thus, the period used within the analyses produced in this thesis, spanning 2005-2009, has a higher proportion of non-recommendations in its sample than observed in other publications reporting SMC decision-making. Dr. A. Walker[4], suggested that such differences can be attributed to the use of different time horizons (2005-2009 vs. 2007-2009 or single years) and also to the fact that publications report SMC coverage decisions across all types of submissions, whether full submissions, resubmissions, abbreviated submissions or IRPs. Within the SMC sample used for this thesis, only full or re-submissions were included.

Multivariate analysis of SMC decision-making suggest that seven variables appear to have a significant effect on coverage decisions: the sample size and duration of the RCT(s), the ICER, if the technology was indicated for use in infections or skin diseases, the prevalence of the disease in question, as well as whether there was an alternative therapy available within NHS Scotland (Table 9.1). An increase in the ICER had a significant positive effect on the odds of a restriction and non-recommendation relative to a recommendation. This result is in line with the stated objective of the SMC which is to recommend for use those technologies representing value for money for NHS Scotland. The fact that budget impact of new technologies was not found to impact on the outcomes within the models described here reinforce the fact that the SMC is focused on promoting 'efficiency' of funding rather than a concern for affordability. The ICER and disease prevalence were the only variables that had significantly

[4] Walker, Andrew. Senior Lecturer in Health Economics, University of Glasgow and member of the New Drugs Committee at the Scottish Medicines Consortium. Interviewee: Karin Cerri. Interviewed by telephone, on December 20th 2010. Meeting minutes are provided in Appendix E.

increased the odds of both a restriction and non-recommendation, relative to a recommendation.

Clinical evidence and disease characteristics were found to be key drivers of SMC decision-making. The odds of non-recommendation decreased with increasing RCT size and duration, reflecting the impact of the clinical evidence base on SMC decision-making. When comparing restricted versus recommended technologies, an increase in the sample size of the RCT appeared to have an effect in the opposite direction: it appeared to increase the log odds of a restriction, and decrease the log odds of a non recommendation. However, only the latter effect was statistically significant. This suggests that use of a multinomial outcome variable facilitates the ability to identify potential differences in impact that would not have been possible with a binary outcome category. In addition to clinical variables, disease characteristics were found to increase the odds of recommendation. Moreover, technologies indicated for the management of infectious diseases or skin diseases also increased the odds of a recommendation.

The presence of an alternative therapy indicated in the same population to the technology under evaluation appears to impact significantly on SMC coverage decisions, although perhaps not in the direction that could have been hypothesised based on the literature. The analysis suggests that the availability of an alternative therapy increases the odds of a recommendation. In other words, those technologies where no alternative was present had increased odds of non-recommendation, relative to those technologies where an alternative was already available in NHS Scotland, all other things being equal. It is noteworthy that about 40% of those technologies with no alternatives were orphan-designated technologies. In addition, those technologies where no alternative was available had higher mean ICERs compared to those technologies where alternatives were available – almost three times as high (£62,021 vs. £20,679 respectively). The fact that technologies for which alternatives were not readily available within the Scottish NHS were more likely to not be recommended suggests that while clinical characteristics and need are recognised criteria within SMC decision-making, the ICER would appear to have a greater impact on coverage decisions. This observation is supported by research of SMC coverage decisions for orphan-designated pharmaceuticals in which the 49% of technologies that were not recommended had higher ICERs than those recommended or restricted (Vegter et al. 2010). While the authors did not adjust for other confounding factors in the analysis,

their descriptive analysis confirms that focus of the SMC on "value for money". In addition, these results raise questions about the extent to which the SMC encourages innovation in the pharmaceutical industry. As highlighted by the two quotes at the beginning of this chapter, there is a balance that needs to be struck between encouraging 'value for money' and patient access to innovative care. These particular results would suggest that the SMC, while fulfilling its objective of recommending the use of technologies that demonstrate value for money, may be achieving this objective at the expense of access to innovative care.

When examining the results of the SMC multivariate analyses, there are several limitations that need to be taken into account. The first is that the SMC, compared with other agencies like NICE, provides relatively limited information in the public domain on the evidence reviewed and considered in its decision-making process. In general, the publicly available SMC Advice reports provide a concise review of key issues in a summarised format that do not document details on the various clinical considerations or economic arguments to which they were exposed or considered, but primarily those considerations which were found to be drivers of their decision. On the one hand, this helps the data extraction process by providing the key data that was felt, by the agency, to drive its decision-making. On the other hand, the aim of this thesis and analysis was to collect as much objective evidence as possible on factors driving decision-making. The lack of detail in reporting led to higher rates of non-reporting of variables of interest for this research, particularly with regard to information on the uncertainty around incremental cost-effectiveness ratios, compared to agencies like NICE where a larger quantity of information is publicly available, including manufacturer submissions (depending on the appraisal process used). The lack of data linked to this reporting style was managed by using imputation techniques in the multivariate analysis. The implications of using such techniques, versus restricting the analysis to complete observations, were assessed in a sensitivity analysis in which the multivariate analysis was conducted on the sample of coverage decisions for which the data was complete (130 of 288 appraisals). It is recognised that excluding incomplete observations may lead to bias in the coverage decisions included in the analysis due to the fact that incomplete observations may be systematically different from complete observations. Despite this potential for selection bias, four of the seven variables that were significant in the base-case analysis remain significant in this sensitivity analysis, confirming their important role in SMC coverage decisions.

An important factor to take into account when examining SMC coverage decisions is the agency's reliance on manufacturer submissions in formulating advice. There is no third party or significant additional new analysis performed on the evidence submitted by the manufacturer. The focus of the SMC is to critically review the submitted evidence in order to ascertain the degree of certainty around the effects and value for money of the technology under appraisal. Given the lack of accessibility to manufacturer submissions in the public domain, it was not possible to take into account in the analyses to what degree the SMC advice was driven by the manufacturer submission strategy relative to SMC decision-making criteria. For example, for a technology which was accepted for restricted use, it was not possible to ascertain if this restriction was proposed and implemented by the SMC, or whether the restriction was proposed by the manufacturer in their submission. While there was a lack of complete information on the evidence and manufacturer strategy used within SMC assessments, the analysis presented in Chapter 5 was able to give insights into the effect and significance of clinical, economic and disease characteristics of the technologies assessed by the SMC on its coverage decisions.

### *9.2.3 CVZ*

> "The [Dutch] government policy aimed at reducing drug expenditure appears to bear fruit. Total drug expenditure increased this year by 2.6 percent to 5.2 million [euros] per year. In 2007 the increase was three times as high." (NOS 2008 p1)

The analysis of the Dutch HTA agency (CVZ) looked at 256 technology appraisals. The most common coverage decision by the CVZ was to recommend new technologies (51%), followed by restriction of funding (33%), while 16% of coverage decisions did not recommend funding the technology. The model of coverage decisions by the CVZ included nine variables, representing a mixture of clinical, economic and process variables (Table 9.1). Unlike the NICE and SMC models, the ICER did not have an impact in the CVZ model. However, budget impact associated with the introduction of the technology had a significant effect: a unit increase in the budget impact increased the log odds of a restriction. This may reflect the fact that the Dutch Reimbursement system aims to reduce the growth in out-patient drug expenditure while maintaining high quality healthcare and patient outcomes (Pronk and Bonsel 2004). The impact of budgetary impact considerations on the odds of a non-recommendation was not

statistically significant, suggesting there are other factors that better explain non-recommendations. Interestingly, in this model, the impact of the therapy area for which the technology was indicated played an important role in coverage decisions. Technologies for cancer were associated with an increased probability of recommendation, while technologies for the treatment of cardiovascular disease, and obstetrics/gynecology/urinary-tract disorders increased the probability of a restriction. These results obtained within the CVZ sample suggest that coverage decision-making is a complex decision process involving numerous clinical, disease and affordability considerations. This fits well with the argument presented by Stolk and Poley (2005) that an understanding of CVZ coverage decisions requires a holistic and comprehensive assessment of multiple factors.

While the base-case model was based on a three-category outcome variable, a sensitivity analysis was conducted to assess the effect of using an alternative classification of the outcome variable, in this case a binary outcome category. The results of this sensitivity analysis suggest that several of the explanatory variables which were important in the CVZ base-case model continued to be significant in this binary model. For example, the use of an active comparator and the demonstration of clinical superiority maintained their effect and significance, suggesting that the role of these variables in explaining CVZ coverage decision-making is robust to changes in model-specification. Other variables, including the budget impact, were no longer found to have a significant effect in this sensitivity analysis, suggesting that the use of binary outcome category does not allow for a more detailed exploration of the impact of the budget impact on different coverage decisions, and thus when examined in this sensitivity analysis its overall impact was not significant. Variables that were not included in the base-case model were found to have a significant role in this sensitivity analysis - the year of appraisal and technologies indicated for the treatment of infectious diseases. This suggests that the use of binary outcome categories can yield an alternative perspective on CVZ decision-making, at the expense of reducing visibility on the impact of explanatory variables on specific types of coverage decisions. It also shows that the majority of factors during CVZ decisions retain their effect on the outcome variable despite the use of a binary outcome variable.

The results obtained within the CVZ multivariate model can also be compared with the results of a discrete choice experiment performed amongst Dutch healthcare

professionals who were asked to select, among 27 choice sets, the technology they would chose to reimburse (Koopmanschap et al. 2010). The choices made were analysed using multinomial logistic regression and suggested that severity of disease, cost-utility analysis, patient outcomes and budget impact were the most significant criteria driving coverage decisions (Koopmanschap et al. 2010). These results have strong similarities with the results of the CVZ multivariate analyses performed within this thesis. Cancer therapies, which could be approximated to represent severe disease, significantly increased the odds of recommendation. Budget impact was also found to play a significant role in CVZ decision-making. An important difference between this analysis of CVZ decision-making and the analysis performed by Koopmanschap et al. (2010) is the effect of the CUA on coverage decisions which was noted in the latter study. In Koopmanschap et al. (2010), results of the discrete choice experiment suggest that CUA is an important criterion for CVZ decision-making. This was not found to be the case in the multivariate analyses performed within the context of this research. Plausible explanations for this difference could be due to variation between hypothetical reimbursement decisions versus real-life decision-making, and the fact that CUA was first introduced in the CVZ process in 2005, and is only utilised as a criterion for inclusion of technologies on List 1B. It has been highlighted that the role of CUA within the CVZ appraisal process may increase in the future (Dr. Graaff, M; Dr. Goettsch, W, Dr. S. Kleijnen)[5].

Interpretation of CVZ multivariate results needs to take limitations into account. An important limitation is related to the ability to access only information that is publicly available. The CVZ only presents in the public domain the information corresponding to the final recommendation. For instance, a technology that is recommended for GVS List 1A, meaning that it is clustered with therapies already available within the Dutch healthcare system, will have the report/information used to support this coverage decision are made publicly available. However, it is possible that a manufacturer may have submitted a request for the drug to be included in GVS List 1B i.e. where the technology is found to have therapeutic benefit to the degree that it is not clustered with existing therapies. This type of submission would have required the technology to demonstrate its cost-effectiveness and thus the manufacturer may have submitted cost-

[5] Dr. Martin van der Graaff, Secretary medicines evaluation committee (CVZ); Dr. Wim Goettsch; Sarah Kleijnen (M.Sc. ) Project coordinator EUnetHTA WP5. Interviewee: Karin Cerri. Interviewed by telephone, on January 6th 2011. Meeting minutes are provided in Appendix E.

effectiveness analyses. However, if the decision by the CVZ/CVZ was that there was no added therapeutic benefit to justify inclusion in the 1B list, then the cost-effectiveness criteria would not have been applied, and therefore the information about the cost-effectiveness analyses that were performed would not be disclosed in the public domain (See Chapter 6).

From one perspective, the lack of access to data that was not actually taken into account in the coverage decision does not have significant implications for the analysis, given that the aim is to identify those factors that impact on decisions. However, this perspective assumes that the availability of data has no impact on the committee's decision to cluster or not cluster the technology. To some degree the implications of incomplete evidence was addressed through imputation techniques, coupled with the use of dummy variables to identify the impact of incomplete observations data on coverage decisions, and an additional sensitivity analysis in which the model was restricted to those technologies with complete observations. This sensitivity analysis highlighted that the majority of observations from the CVZ were incomplete (96 of 256 were included in the analysis). This suggests that there is considerable additional evidence and data provided to the CVZ that is not disclosed (e.g. patient submissions, physician organisation interaction, cost-effectiveness analyses etc). Therefore, greater transparency in the evidence received or submitted may help further increase the understanding of CVZ decision-making.

As with the SMC appraisal process, the CVZ also relies on manufacturer submissions in formulating its advice. There is no third party or significant additional new analysis performed on the evidence submitted by the manufacturer. Given the lack of accessibility to manufacturer submissions in the public domain, it was not possible to take into account in the analyses to what degree CVZ recommendations are driven by the manufacturer submission strategy relative to CVZ decision-making criteria. For example, for a technology which was accepted in the GVS 1A list (clustered technologies), it was not possible to ascertain if the inclusion in this list was proposed by the CVZ, or by the manufacturer in their submission. Despite the lack of access to such information, it does not detract from the possibility of being able to assess the degree to which key characteristics and factors vary according to the coverage decision made.

### *9.2.4 HAS*

> "The ASMR obtained reflects the recognition of quantitative improvement (the efficacy of A is twice the efficacy of B), a qualitative improvement (A treats patients not responding to B), and/or a tolerability improvement, compliance improvement, or improvement in the therapeutic maintenance of effect over time..." (Bouvenot 2006 p. 11) [1]

The analysis of HAS decision-making focused on understanding the factors driving ASMR ratings. These ratings reflect the perceived incremental medical value associated with the appraised technology relative to standard of care, and as highlighted by the quote above, the ASMR rating can reflect a variety of forms of therapeutic benefit. In exploring the factors driving coverage decisions made by HAS, 315 technology appraisals were reviewed: 3% of technologies were awarded an ASMR I, meaning that the technology was considered to bring highly significant incremental medical benefit and 15% of technologies were awarded an ASMR II, in instances where the committee considered that the technology would bring important incremental medical improvement. The majority of decisions (44%) concluded that there was no medical improvement associated with the technology (ASMR V). When HAS coverage decisions were modelled, nine clinical, disease and socio-economic variables appeared to have a statistically significant impact on the odds of ASMR III-IV or ASMR V relative to ASMR I-II. No economic variables were included as the HAS does not include economic criteria in its appraisal process.

Clinical and disease characteristic variables played an important role in HAS decision-making in this model (Table 9.1). Technologies that demonstrated clinical superiority increased the log odds of an ASMR I-II. This could be seen to reflect the focus of the HAS appraisal process on the incremental medical benefit associated with the technology, demonstration of clinical superiority being a key piece of evidence to substantiate the presence of incremental medical benefit. It is noteworthy, however, that for those technologies that demonstrated clinical superiority, this superiority was demonstrated only 25% of the time versus an active comparator. For those technologies that did not demonstrate clinical superiority, 68% were compared to active

---

[1]Translated from the following original quotation: "L'obtention d'une ASMR peut traduire la reconnaissance soit d'un progrès quantitatif (A est deux fois plus efficace que B) soit d'un progrès qualitatif (A permet d'atteindre des patients non-répondeurs ou insuffisamment répondeurs à B) ou traduire une meilleure tolérance, une meilleure observance ou un meilleur taux de maintenance thérapeutique…"

comparators. Therefore, these results would suggest that with regard to HAS decision-making, the demonstration of superiority in itself outranks the nature of that superiority (whether demonstrated versus placebo or an active treatment). This may also suggest that the criteria prioritized by HAS in its assessment of medical value matches closely with that used by regulatory agencies that also accept, and indeed in several circumstances recommend, comparison to placebo as an appropriate means of demonstrating clinical benefit.

When examining disease characteristics and their impact on HAS decision-making, technologies that had an orphan designation increased the log odds of an ASMR I-II. This could be explained by the fact that orphan-designated technologies tend to be rare, indicated for diseases with no alternatives. When the features of orphan technologies were further examined, they were found to be characterised by small patient populations (mean estimated target population for orphan technologies was 1,356 patients vs. 570,408 patients for non-orphan technologies). In addition, orphan technologies were indicated for diseases with low availability of alternative therapy (in 38% of cases alternative technologies were available) and therefore a correspondingly higher level of clinical need for treatment. For non-orphan technologies, in 80% of cases alternative therapies were available within the French healthcare system. Orphan designation increased the odds of high ASMR ratings despite the fact that, compared to non-orphan technologies, orphan technologies on average were supported by fewer RCTs (1.5 vs. 2.8), had mean shorter trial duration (29 weeks vs. 53 weeks), were supported by smaller trials with, on average, fewer patients (255 patients vs. 1,228 patients); and had a higher proportion of instances in which an active comparator within the clinical trial was not available (34% vs. 16%). This evidence would support the hypothesis that the HAS, in its assessment of orphan technologies, may be willing to place more emphasis on the potential for the technology to fill a specific clinical need, at the expense of the quality of clinical evidence.

The impact of specific disease areas on coverage decisions was also examined and found to be important within HAS decision-making. Technologies that had a license for the treatment of musculoskeletal and joint diseases increased the log odds of an ASMR I-II. In contrast, indications for the treatment of CNS disorders and infectious diseases increased the probability of an ASMR III-IV and ASMR V. This could be explained by the fact that a higher proportion of technologies indicated for the treatment of

musculoskeletal and joint diseases were supported by clinical trials with active comparator arms (55%). Use of active comparators in clinical trials confers more useful evidence to ascertain the incremental clinical benefit of a technology to standard care. Placebo-controlled trials are less useful in that they provide evidence of incremental benefit of a technology relative to a comparator which does not exist in clinical practice (i.e. placebo is not used to treat patients). In contrast, technologies for infectious disease or CNS disorders had a lower proportion of studies with active comparators (36% and 22%, respectively). Overall, the multivariate model emphasised the role of key clinical and disease criteria on ASMR ratings from the HAS.

In terms of socio-economic factors, pharmaceutical expenditure was found to have a statistically significant impact on the log-odds of both ASMR III-IV and V. A unit increase in pharmaceutical expenditure, in this case national average per patient expenditure per year, appeared to increase the odds of an ASMR III-IV or ASMR V, albeit statistically significant only in the latter. This would seem counter-intuitive in that the increase in pharmaceutical spending would generally suggest an increase in the available budget for reimbursed technologies. However, it could be that an observed increase in pharmaceutical expenditure may have triggered more stringent assessment of incremental medical value. This is conjecture that cannot be further examined with this dataset, and highlights that caution is needed in the interpretation of the role of socio-economic factors such as national pharmaceutical expenditure as there are numerous unmeasured factors that could be associated with this particular variable (e.g. overall trends in GDP, change in treatment algorithms, healthcare system approach, physician and patient behaviour, industrial policy, increases in marketing authorisations for pharmaceuticals etc).

When examining the results of the HAS multivariate analyses, there are several limitations that need to be taken into account. The first is that HAS, compared with NICE for example, provides relatively limited information in the public domain on the evidence reviewed and considered in its decision-making process. In general, HAS reports made publicly available provide concise synopses of key issues in a summarised format that do not document details on the various clinical considerations or disease characteristics which were considered. The lack of detail in reporting did lead to instances of non-reporting of variables of interest for this research. The lack of data linked to this reporting style was managed by using imputation techniques in the

multivariate analysis. The implications of using such techniques, versus restricting the analysis to complete observations were assessed in a sub-analysis in which the multivariate analysis was conducted on the sample of coverage decisions for which the dataset was complete. The results of these sensitivity analyses confirmed the important impact of the majority of indentified variables in the base case analysis.

Another important factor to take into account when examining HAS coverage decisions is that the sample of appraisals included in the analysis was, in fact, a subset of the total pool of appraisals conducted by the Agency. This sub-sample included technologies that also had been appraised by NICE and the SMC in 2004-2009. All the HAS recommendations linked to these technologies were extracted for review. The rationale for this approach was the fact that HAS has numerous responsibilities, one of which is the provision of advice on new technologies available for patients, and in total, its Transparency Commission issued more than 2600 recommendations related to medications in 2004-2009 (HAS 2009). Given the resource constraints available, it was not possible to review all 2600 recommendations to identify those of relevance for this research (i.e. not all recommendations provide ASMR, some recommendations are related to new mode of administration, new safety information or a re-review of technologies licensed prior to 2004). The benefit of the approach employed here, i.e. in taking a sub-sample, was that it increased the opportunity for comparability across agencies by collecting information on a common list of technologies, and secondly it facilitated the streamlining of data extraction to those appraisals of relevance for the research question.

A further limitation associated with the HAS analyses is that this HTA body focuses on determining an ASMR rating. Therefore, HAS does not provide information on the degree to which the French healthcare system is willing to pay for incremental medical benefit (as defined by the ASMR rating). This is directly linked to the role of the HAS which is focused on evaluating the medical benefit of the technology. The output of this assessment is then used by a separate organisation (CEPS) to negotiate a final price. Thus, understanding HAS decision-making may not provide a full perspective on coverage decisions within France and how public funding is allocated to pharmaceuticals. On the other hand, it is recognised within French legislation that the degree of medical benefit (as defined by the ASMR) directly impacts on the price of a technology. Technologies with no incremental benefit are not included on the

reimbursement list unless they are discounted below the level of the other available treatments already reimbursed by the system. Technologies with an ASMR IV can obtain, at a maximum, a price that is equal to the comparator that is already reimbursed while it is only those technologies with an ASMR I-III that can aspire to potential premium prices. Since details of pricing negotiations and discounted prices are not in the public domain this prevented the inclusion of an economic component in the HAS analyses. Thus, while the multivariate analyses of HAS decisions presented here cannot directly examine the economic value that the French system attaches to particular degrees of medical benefit, it does provide an indirect view, by examining the factors that drive HAS allocation of ASMR ratings to the technologies it assesses.

### *9.2.5 Pooled analysis*

In addition to modelling coverage decision-making for each agency separately, a pooled analysis of all coverage decisions across HTA bodies was conducted. An important objective of the pooled analysis was to describe the characteristics of the pooled data set and identify differences between HTA bodies in the nature of technologies appraised and the clinical evidence considered, the process through which HAS appraised the technologies and the socio-economic context in which the appraisals took place. In addition, an important objective was to test whether an "HTA body effect" on the odds of recommendation, restriction or non-recommendation while adjusting for a range of confounding factors. Before embarking on a pooled analysis of coverage decisions across HTA bodies, the pros and cons of such an analysis were assessed. Firstly, in favour of conducting a pooled analysis was that it could provide data to help explain variation in coverage decisions across HTA bodies, which is of direct relevance to the research question. Secondly, the pooling together of appraisals from four HTA agencies could significantly increase the sample size to 977 appraisals, creating the largest single set of data in Europe on HTA coverage decisions and accompanying appraisal characteristics. Pooling across HTA bodies was felt to be feasible due to the fact that the data set was created specifically for this research project, and all data were extracted by the same researcher with a specific data extraction protocol that increased consistency in how the data was extracted.

Overall, the results of the quantitative analysis of pooled coverage decisions made by NICE, SMC, CVZ and HAS was useful to demonstrate that a combination of evidence, process and socio-economic context factors help explain variability in coverage

decisions made across these four HTA bodies. The analyses also represent the first model of coverage decisions across four European HTA bodies based on 30+ explanatory variables extracted directly from their appraisals. This analysis suggests that those factors of importance in explaining variation in coverage decisions across HTA bodies may not be the same as those factors explaining variation in coverage decisions within each HTA body – namely, process and socio-economic context factors play a more evident role in this analysis than in the single HTA analyses. At the same time, the impact of the clinical and disease factors observed in the single analysis remains very significant in the pooled analysis.

While the results of the pooled analysis were felt to be meaningful and helpful in increasing the understanding of drivers of coverage decisions, it is important to recognise the important limitations that must be taken into account when interpreting the results of this pooled analysis. Firstly, a significant challenge in conducting pooled analyses of coverage decisions is that across the four HTA bodies analysed, coverage decision are not formulated in the same way, and that heterogeneity was observed within the 'restricted' category in how HTA bodies implemented restrictions of coverage. To evaluate the robustness of the base case model, a sensitivity analysis was implemented using a binary outcome variable; this provided very similar results to the base-case pooled analysis.

A second important challenge identified in implementing a pooled analysis of coverage decisions is to create as homogenous a platform as possible upon which to model. To overcome the potential limitations associated with this, in the base case model, the heterogeneity of the samples across HTA bodies was adjusted for by including in the model variables that captured the nature of the disease corresponding to the technologies appraised. In addition, two sensitivity analyses were conducted in a sub-set of the total sample that was hypothesised to increase the homogeneity of the technologies assessed between HTA bodies. Overall, the results of these sensitivity analyses show that the explanatory power of the combination of clinical, economic, process and socio-economic factors is higher in the sub-analysis including only those technologies appraised by all four HTA bodies. This may suggest that increasing the homogeneity of the sample facilitates a more robust analysis of the role of explanatory variables on coverage decisions. On the other hand, homogeneity is obtained at the

expense of significantly restricting the sample of analysis to less than 20% of the total available sample, increasing the odds of selection bias in the sample.

Overall, the pooled analyses provide new evidence that differences exist between HTA bodies in the nature of technologies appraised and the clinical evidence considered, the process through which HTA bodies appraised the technologies and the socio-economic context in which the appraisals took place. Importantly, the evidence also shows that the HTA body effect is present which contributes significantly to decision-making, even when adjusting for a range of confounding variables. The analyses also emphasise that, amongst those HTA bodies that consider cost-effectiveness analyses, the ICER has a significant effect on the coverage decision. The pooled analysis has alos provided an opportunity to examine the role of process and socio-economic factors on decision-making.

### 9.3 General limitations

In addition to specific limitations and challenges encountered in the individual and pooled HTA analyses, there are some overarching limitations that apply more generally. The selection of HTA agencies included in the analysis was in part limited by data availability. In particular, only those HTA agencies that put their decision-making process and reports into the public domain were considered for analysis. Thus, extrapolating to other HTA bodies not included in this research is not advisable without having a more concrete understanding of their decision-making processes. In addition, the generalisation of results, even within the respective agencies, must be attempted cautiously. HTA agencies are continuously evolving and factors that may have driven their decision-making during 2004-2009 may not be the same or remain constant over time. For example, it was highlighted in an interview with CVZ representatives (Dr. Graaff, M; Dr. Goettsch, W, Dr. S. Kleijnen)[2], that the role of cost-effectiveness evidence may become more prominent within CVZ decision-making in the future. The Conservative-Liberal Democrat government elected in the UK in May 2010 has indicated that they will change NICE's role in England and Wales such that its reimbursement recommendations will only have advisory status. While this change in NICE's role may or may not impact on the factors driving its decision-making, these

[2] Dr. Martin van der Graaff, Secretary medicines evaluation committee (CVZ); Dr. Wim Goettsch; Sarah Kleijnen (M.Sc. ) Project coordinator EUnetHTA WP5. Interviewee: Karin Cerri. Interviewed by telephone, on January 6th 2011. Meeting minutes are provided in Appendix E.

changes would appear to impact most significantly on the implementation of NICE guidance by the health services (further discussed in section 9.5.2).

While every effort was made to create a dataset and analysis that was comprehensive in its inclusion of explanatory variables, it was not feasible to capture all variables presumed to have a potential impact on coverage decisions. Although the analysis captured information on the ICER and the uncertainty around it, these indicators do not provide information on the nature of the economic model, the design, comparators and subtleties of the analysis. Moreover, the use of surrogate outcomes was not captured as a specific variable in the dataset: there is growing interest in better understanding the use of surrogate endpoints in decision-making (Velasco Garrido et al. 2009). As the literature does suggest that most clinical trials utilise surrogate endpoints, it was considered impractical to include this as a variable due to potential lack of variability between technologies, and other variables linked to the clinical characteristics of the technology were prioritised instead (such as use of active comparator in trial and demonstration of superiority in clinical trial).

The speed of the appraisal process was also not factored in the analysis. The fact that the SMC takes 3 months and NICE takes 12 months to complete an appraisal could have an impact on the nature of the evidence considered and the impact of the process on outcomes. The analysis is also limited in that it cannot systematically examine time-to-coverage decisions as a specific outcome of analysis. Finally, the data base excluded those appraisals for which no documentation was available. This includes situations where manufacturers did not make a submission, and therefore no appraisal was conducted. This could be considered as a bias in the sample in that it did not consider the characteristics of those technologies that were not appraised. However, as information was not publicly available, such analyses could not be conducted. In general, the proportion of such cases excluded due to non-submission was relatively low (e.g. four non-submissions in the NICE dataset in 2004-2009). Within the SMC appraisal process, if non-submission takes place, the technology receives a 'non recommendation'. Thus, the exclusion of these non-submitted appraisals may lead to an under-estimate in the sample of the true proportion of non-recommendations. However, the level of non-submission within the SMC was also low.

Socio-economic indicators, such as GDP, are known to be influenced by many different factors. Indeed, such indicators act as a surrogate for many characteristics of the country it applies to. In addition, such indicators that vary at the HTA body level, rather than the technology appraisal level, are unlikely to have a very strong effect, due to the limited number of HTA bodies in this analysis. Therefore, the interpretation of the impact of such broad indicators, such as the percentage of GDP spent on healthcare, will need to take into consideration the risk that variations observed in such indicators across HTA bodies may be correlated with other factors.

Finally, it is important to consider the research findings within the context of the respective origins and objectives of NICE, SMC, CVZ and HAS. The evolution of the HTA bodies, their origins and their roles within pharmaceutical regulation and coverage decisions vary. In particular, their roles within their respective healthcare systems are very much driven by the context and characteristics of the healthcare system, and each HTA agency has peculiarities in their roles linked directly to that system. HTA bodies, both within and outside of Europe, vary in their objectives, and in the approach and methods used to implement HTA within their jurisdictions (Neumann et al. 2010). The scope of this thesis was limited to national level HTA bodies. In addition, the analysis of factors influencing coverage decisions were performed on decisions made within HTA processes. However, within a healthcare system, where funding/reimbursement decisions are confirmed at a national level, regional and local payers may re-assess whether funding/reimbursement should be provided. Due to time constraints, factors influencing regional/local funding/reimbursement decisions were not assessed.

### 9.4 How this thesis contributes to the literature

The literature has made important contributions to assessing the role of evidence, process and context factors on coverage decision making. However, limitations were identified in the literature in terms of scope, analytical methodology and comprehensiveness that hampered the ability to robustly examine how HTA agencies address the paradox of coverage decision-making, and the balance between evidence, process and context factors that drive decision-making. In relation to these gaps, this thesis provides new evidence that contributes in a number of ways: the inclusion and comparison of multiple European HTA bodies, a comprehensive analysis of multiple factors through the creation of a bespoke database, and the adoption of multivariate

analyses to identify those factors that drive coverage decisions within each HTA body and across HTA bodies.

This research includes four different HTA agencies from different countries, thus broadening the scope of evidence available on factors driving European HTA coverage decisions. Much of the currently available literature on European HTA bodies currently available focuses on NICE as an example of HTA in Europe, despite the fact that in recent years there has been a growth in the use of HTA across EU Member States (OECD, 2005). In addition, where cross-agency analyses have been performed, these included Anglophone agencies, namely NICE and SMC, alongside CDR in Canada and PBAC in Australia. Moreover, where multiple European HTA agencies have been compared and contrasted, this literature has examined the characteristics and differences in processes between HTA bodies, but without linking such differences in characteristics and processes to explain variation in coverage decisions (Sorenson et al. 2008; Hutton et al. 2006; Kanavos et al. 2010). In providing a fresh perspective on factors driving HTA decisions within four distinct HTA bodies, this thesis contributes to and augments the existing evidence base in this area of research.

In addition to adopting a broader scope to provide new evidence on HTA decision-making in Europe, this thesis was designed to contribute to the understanding of HTA coverage decisions in Europe by considering a broad range of evidence, process and context variables in understanding coverage decisions. The current literature that examines factors driving decision-making has, to a large extent, focused on specific types of factors (e.g. economic factors or process factors). Few have combined factors (e.g. Clement et al. 2009; Dakin et al. 2006; Devlin and Parkin 2004), and there are no studies which have aimed to compare HTA decision-making using a dataset that combines evidence, process, and socio-economic context variables extracted directly from HTA reports and related sources. While process and context-related factors have been identified as potentially important influencers of coverage decisions, and differences in these factors between HTA bodies have been described, few authors have in fact made a link between differences in process and context factors within coverage decisions. Thus, this thesis contributes to the literature by capturing within a single analysis a comprehensive range of evidence, process and context factors.

A key contribution of this thesis is the use of a bespoke dataset to capture information on a broad range of factors thought to impact on the decision process, and which has the ability to observe the interaction of these factors and their relative contribution to coverage decision-making relative to one another. In the literature, an important gap was observed in terms of the analytical methods adopted to assess coverage decisions and relevant factors, and the ability to distinguish the role of one factor while adjusting for the presence of other important characteristics. Important exceptions are the models of coverage decisions that have been developed for NICE (Dakin et al. 2006; Devlin and Parkin 2004). With respect to other HTA bodies, neither single-agency analyses nor comparative analyses across several European HTA bodies have been conducted to robustly assess the relative contribution of a range of factors on coverage-decisions. Comparative analyses across HTA bodies were primarily qualitative or have adopted descriptive quantitative methodologies (Barbieri et al. 2009; Clement et al. 2009; Kanavos et al. 2010; Vegter et al. 2010). Such descriptive analytical techniques make it difficult to interpret the relative contribution of each factor, given the absence of adjustment for other factors in the analysis. In response, this research created a bespoke dataset of HTA coverage decisions from four HTA bodies over a five-year period and utilised statistical methods of analysis to assess the relative contribution of a comprehensive range of factors on coverage decisions both within and across HTA bodies.

## 9.5 Implications of this research

### *9.5.1 Heterogeneity between HTA bodies - implications for healthcare systems and patients*

The thesis has shown that British, Dutch and French HTA bodies differ in the pattern of coverage decisions, and each is governed by a specific mix of evidence, process and context factors that drive decision-making. The factors that best explain the variation in coverage decisions for NICE are not the same as those that explain the variation in coverage decisions for SMC, CVZ or HAS. This research confirms through quantitative analysis that there is a difference between HTA bodies in the pattern of coverage decision made, and key differences in the factors that drive decision-making.

From one perspective, such differences in coverage patterns and the factors that drive those coverage decisions can be explained by the fact that each HTA body is designed to match as closely as possible the specific healthcare system it serves. Therefore,

variation observed across HTA bodies could be considered to reflect the reality of the healthcare market in Europe; that is, HTA bodies are actors within specific healthcare systems, operating within the context of regulatory systems that combine "... public health, healthcare and industrial policy interests, reflecting particular national circumstances and requirements" (Permanand 2006 p. 5). Unlike other functions which have been centralised at the European Union level, such as regulatory processes and marketing authorisations, the subsidiarity principle allows for the independent development of EU Member State healthcare systems to continue. From this point of view, acceptance of autonomy and variation in healthcare systems in Europe should imply acceptance of variation in HTA coverage decisions and the factors driving such decisions.

However, from another perspective, differences in the proportion of recommendations, restrictions and non-recommendations can be seen to be contrary to the principle of equitable access to treatment in Europe (Wilking and Jonsson 2005). This is due to the fact that coverage decisions impact on the ability of patients to access pharmaceuticals, and that variation in coverage decisions are then linked to variation in patient access to pharmaceuticals. Proponents of equitable access to pharmaceuticals have argued in the literature, the media and through patient and physician lobby groups, that variation in access to pharmaceuticals is unacceptable for patients:

> "…differences in access to new innovative oncology drugs cannot persist: cancer patients in Europe will not accept that a standard of care available in one European country is not available in other countries" (Wilking and Jonsson 2005 p. 94)

This quotation highlights the importance given to consistency of access to pharmaceuticals as a key concern for patients in Europe. From this perspective, the notion of one HTA body recommending and another HTA body not recommending the same technology for use is incongruent with the belief that patients living with the same condition in different European Member States should have the possibility to access the same pharmaceuticals. Research findings within this thesis confirm that HTA bodies diverge significantly in the coverage decisions made for the same set of technologies. Such differences between HTA bodies were observed while adjusting for a range of clinical, disease, economic, process and socio-economic factors within a sample of technologies that was common to all four bodies.

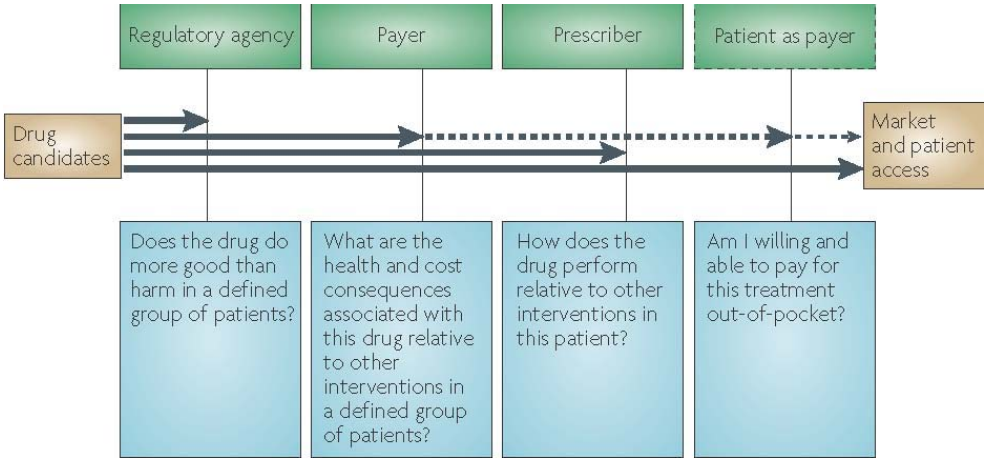
While the decisions of HTA bodies may impact differently, some more favourably than others, on the possibility of patients accessing pharmaceuticals, there is also evidence to suggest that HTA bodies are sensitive to the issue of patient access, perhaps contrary to some media reports. Descriptive and multivariate analyses found that orphan technologies were more likely to be in the recommended group rather than the restricted or not recommended group by NICE, HAS and CVZ. Treatment of severe diseases, in particular cancer, were found to be more likely to increase the probability of recommendation rather than non recommendation in NICE, CVZ and HAS decision-making. This increase in probability of recommendation was observed across the pooled sample of technologies and despite the fact that both orphan-designated technologies and cancer therapies tended to have a weaker evidence base and less certainty around the cost-effectiveness profile. It is noteworthy, however, that the impact of orphan status or cancer therapy was not observed within the SMC sample. In contrast to the other three HTA bodies, the odds of recommendation tended to decrease for technologies without any alternative available within NHS Scotland, of which 40% were orphan designated technologies.

It is also important to recognise that HTA bodies are not the sole determinants of access to pharmaceuticals in Europe. Affordability at the level of the provider is an important factor in determining patient access to pharmaceuticals. The decentralisation of healthcare systems in many European countries has meant that affordability decisions are increasingly made at regional and local levels. The process of decentralisation has taken on multiple forms across Europe from the devolution of decision-making powers to privatization, driven by healthcare systems' specific needs and structures (Saltman et al. 2007). Such variation in the definitions and concepts of decentralisation, and variation in implementation, impact on local decision-making, and lead to diversity in the funding and availability of technologies (Delamothe 2008). A specific example related to the implementation of NICE schizophrenia guidance suggests that implementation is variable between Primary Health Trusts (PCTs) in England (Mears et al. 2008), which the authors attribute to differences in commitment and the leadership of executive staff within these Trusts. McGuire and Litt (2003) highlight that in addition to the aggregate health-care budget available to the healthcare system, local incentives play an important part in the degree to which efficient allocation of health-care budget is achieved. A review of expenditure on cancer treatment by cancer patients across PCTs shows large variation between regions (£17,028 in Nottingham City Primary Care Trust

to £5,182 in Oxfordshire PCT) (Gubb 2008). Budget constraints at local level mean that trade-offs need to be made by local-decision bodies on which technologies to fund (Leatherman and Sutherland 2008). A recent analysis of variation in access to healthcare in England highlights that there are multiple supply- and demand-side factors that can impact on the degree of variation in access to healthcare (Appleby et al. 2011). Variation in demand for healthcare is driven by variation in a number of factors including commissioning priorities, morbidity, patient choices, GP decisions (Appleby et al. 2011). Variation in the supply of healthcare can also lead to variation in access to healthcare: such variation in supply can arise from government policy, availability of resources, service configuration, and clinical decisions (Appleby et al. 2011).

Thus, while there may be national HTA guidance on the use of a technology, the same technology may not be incorporated into all formularies across the country. This highlights the usefulness of distinguishing between access as the *availability* of a specific treatment versus defining access as actual equality in *utilisation*. Certainly, HTA decisions that do not recommend a technology for public funding will limit the availability of that technology, which is perhaps why negative recommendations receive significant attention. However, it is important to note that positive HTA recommendations may not necessarily translate into patient access. While a positive recommendation makes a particular technology formally available to the healthcare system, the autonomy of member state healthcare systems, decentralisation, local affordability, and professional support for guidance play an important role in determining patient access. Figure 9.1 outlines the various forces that influence patient access to pharmaceuticals (Eichler et al. 2010). Therefore, the findings of this research, and the role of HTA within the wider debate on patient access to pharmaceuticals, need to be considered in the light of the broader healthcare systems and the public funding processes.

**Figure 9.1 Decision makers on the road to market access**




*Source:* Eichler et al. 2010 p.278

### *9.5.2 Implications for the pharmaceutical sector: reward for innovation?*

> "Pharmaceutical innovation requires novelty of effectiveness. Pharmaceutical innovations create value to society by making it possible to generate improvements in patient health (net of treatment risks) that were previously unattainable" (Morgan et al. 2008 p.4)

In the literature on innovation in pharmaceutical care, and as highlighted in the quote above, innovation is linked to the fulfilment of an unmet medical need. The degree of innovation can be defined according to the degree to which it improves health outcomes or meets significant unmet clinical needs (Morgan et al. 2008; Sermet 2007; Achilladelis and Antonakis 2001; De Cock 2010). The larger the medical need that is fulfilled, the larger the innovation (Morgan et al. 2008). Innovation is recognised as a cyclical process, through which future innovations are dependent on current innovations gaining access to healthcare systems, and ultimately to patients, thus providing pharmaceutical companies with revenue to invest in future innovations (Attridge 2006). Innovation can be seen to benefit not only the pharmaceutical industry, but ultimately also patients, healthcare systems, and generics manufacturers (Teece 1987, in Attridge 2006).

The degree to which the HTA bodies examined in this thesis reward innovation can be observed by considering the coverage decisions made for orphan-designated technologies and/or technologies for diseases for which there is no alternative therapy. Based on the definitions of innovation presented above, such technologies would represent significant innovation either by providing treatment for conditions that were

previously not treated, or for providing treatments that impact on health outcomes for patients with severe and rare diseases. In this regard, the first finding of this research is that orphan designation and technologies for which there is no alternative tend to increase the likelihood of coverage by NICE, CVZ and HAS. Technologies with a lack of alternative or orphan designation were more frequently found in the recommended or restricted groups. However, this was not observed within the SMC sample. The results of the analyses performed on SMC coverage decisions suggest that for technologies developed for disease areas which are rare and/or for which no specific treatment is available in the healthcare system prior to the technology under assessment, a higher proportion are not recommended for use in the Scottish NHS. When the incremental cost-effectiveness ratio was examined for these particular technologies, it was clear that the mean ICER for those technologies without alternatives available in the NHS was almost three times higher than the ICERs for technologies with alternatives. Similar results were observed in an analysis of SMC coverage of orphan-designated technologies (Vegter et al. 2010). The analyses conducted in this thesis would suggest that introducing a newly licensed technology in an indication for which other technologies already exist would have a higher chance of recommendation than a new technology for which no alternatives exist. Such technologies are generally referred to as 'me-too' technologies, and represent new technologies that belong to a class of technologies already available to the healthcare system for the same patient population. In the NICE, CVZ and HAS samples of coverage decisions, this trend was not observed.

The changes currently proposed by the UK coalition government for NICE are in line with strengthening a signal to the pharmaceutical industry that innovative medicines will be rewarded. It has set out plans to implement a Value-Based Pricing (VBP) system that will replace the current pricing and reimbursement system. The VBP system aims to "…improve NHS patients' access to effective and innovative drugs by ensuring they are available at a price that reflects the value they bring" (DH 2010 p. 6). This change is set to occur in 2014, and until then, technologies will be appraised via the current mechanisms. The DH is currently reviewing the output of the consultation process which has recently closed, and therefore all available details of this new system are liable to change.

Under the VBP system as currently outlined in the DH (2010) consultation document, the level of value of medicines will be defined through a variety of criteria. Criteria that

will play a role in determining the value of a technology will include not only health-related benefits but also societal benefits, and the degree to which the technology targets areas of significant unmet medical need. In addition, the new system aims to incentivize innovation within the pharmaceutical industry (DH 2010).

The VBP system, as currently described, will impact directly on NICE. Firstly, under this new system, the implementation of NICE guidance will no longer be enforced by law. Affordability will be determined at the local level rather than by NICE at the national level. Secondly, given the broader scope through which value will be assessed may lead to changes in the range of factors considered by NICE in its technology appraisals so as to "reflect all the components that contribute to a treatment's full impact on health and quality of life" (DH 2010 p.10). Thirdly, a key aspect driving current NICE decision-making - cost-effectiveness - is likely to evolve as a concept used within NICE decision-making and NHS pricing. As currently set out in the consultation document, it is proposed that multiple cost-effectiveness thresholds would be made available within the VBP system to reflect the incremental value of the technology in terms of its innovation, its ability to address a high unmet need, and its ability to generate benefits to the wider society (DH 2010). A base-case cost-effectiveness threshold would be defined, representing the maximum value that the NHS would pay for technologies that are associated with no incremental value. The cost-effectiveness threshold would then be adjusted according to the degree to which the technology could provide incremental value in the various domains. To some degree it can be argued that NICE is already applying differential cost-effectiveness thresholds: a descriptive analysis of NICE decision-making found that while the ICER was an important factor in decision-making, exceptions were observed where technologies with above-threshold ICERs were recommended or vice-versa (see Chapter 4). In addition, in 2009 NICE included in its methodology guidelines specific advice to the appraisal committee when appraising so called 'end-of-life' technologies that may be life-extending (NICE 2009). Under the VBP system however, the presence and use of these proposed thresholds may become more explicit. Whether considering the future implications of the public funding of pharmaceuticals in the UK, or other European Member States, from a pharmaceutical manufacturer perspective, the implications of the research in this thesis suggest that coverage decisions can provide signals to the pharmaceutical industry that either stimulate or dampen innovation.

### *9.5.3 Implications for harmonizing the evidence needs of regulatory and reimbursement agencies*

Across the HTA bodies examined within this thesis, a significant proportion of coverage decisions in 2004-2009 did not recommend the use of particular technologies within their healthcare systems. Non recommendation was observed in 14% and 16% of NICE and CVZ coverage decisions, and 44% and 46% of HAS[3] and SMC coverage decisions, respectively. These technologies were not recommended for reimbursement despite obtaining a license for use within Europe from regulatory agencies, primarily the European Medicines Agency (EMA). This would suggest the presence of significant inconsistency between regulatory agencies such as the EMA and HTA bodies in the definition of value and the criteria used to determine the value of technologies. It is well known that regulatory bodies and payers are driven by different roles and objectives. On the one hand, regulatory bodies aim to examine the benefit-risk profile of a technology to ensure that the benefits outweigh safety risks (Eichler et al. 2010). On the other hand, payers, broadly speaking, assess whether a new technology represents value-for-money by considering the degree to which the technology provides incremental benefits over an existing therapy. While a placebo-controlled study may be suitable and efficient trial design for regulatory purposes, it is less able to provide evidence to facilitate comparison of new technologies with existing treatment options (Haynes 1999). It also does not generate relevant information for physicians given that the choice made in clinical practice does not usually use of using a placebo. In addition, regulatory trial design often relies on surrogate endpoints rather than 'hard' outcomes to demonstrate efficacy – while efficient from a development cost perspective and to illustrate efficacy, surrogate endpoints do not provide evidence of the effectiveness of the technology to achieve incremental medical benefit when measured using a hard endpoint or outcome.

When the clinical, economic and disease characteristics of the technologies appraised by NICE, SMC, CVZ and HAS were examined, the descriptive and multivariate analyses suggested that, compared to recommended or restricted technologies, non recommended technologies tended to have a weaker evidence base, higher ICERs, and specific disease characteristics compared with recommended or restricted technologies. Non recommended technologies also tended to be indicated for populations with higher

[3] It should be noted that within the HAS analysis, technologies with an ASMR V were considered as not recommended for the purposes of the pooled analyses.

target population size and, in relation to this, higher estimated drug budget impact. In terms of disease characteristics, non recommended technologies tended to be for diseases where alternatives where available, with the exception of the SMC where the opposite was observed. In general, non recommended technologies tended to have a smaller evidence base (fewer trials, shorter duration, smaller sample size), were less likely to have an active comparator within their clinical trial programme and were also less likely to demonstrate superiority of efficacy within their trial. The limitations of RCTs are encapsulated in an example provided by Herland et al. (2005) who examined the proportion of patients in a GP and specialist practice that match the selection criteria in an asthma-related clinical trial. Of the total pool of GP and specialist practice asthma patients, less than 4% met the RCT inclusion criteria. Such obstacles have led researchers to conduct observational studies to supplement RCT analyses. For example, Cazzola et al. (2010) identified the need to assess the effectiveness of an asthma therapy (omalizumab) in a real-life clinical practice setting due to the fact that the clinical trial data included patients that did not represent the majority of patients in their clinical practice, and secondly the trial data did not provide sufficient information to be able to identify those patients that would be most likely to benefit from the treatment. In this thesis, as part of the analysis of evidence considered by HTA bodies in their appraisal process, information was gathered on the use of observational data, defined as non-randomised non-interventional study designs. Across the four HTA bodies, the use of such evidence was very rare (less than 5% of coverage decisions). This may be primarily due to the fact that in most cases the technology is appraised upon entry to the healthcare system, and as such has not yet had sufficient time to collect real-life clinical data on its use and effectiveness. These findings suggest that the evidence needed to obtain regulatory approval and the evidence needed to obtain HTA recommendation may not be fully aligned.

### *9.5.4 Implications for a harmonised European HTA system*

The European Network for Health Technology Assessment (EUnetHTA) was created to increase European-wide co-operation between HTA agencies, with the objectives of increasing efficiency in the implementation of HTA (Kristensen et al. 2009). One of the chosen means for supporting an increased efficiency in cooperation between HTA bodies has been through the development of tools to support health technology assessments in Europe (Kristensen et al. 2009). The "Joint Action on HTA" programme launched in 2010 by the European Commission and implemented by EUnetHTA, aims

to focus, as a first step, on standardising the methodologies used by HTA agencies in Europe. A pilot project examining harmonisation of HTA for orphan designated technologies is currently underway. In addition, the European Parliament (2011) recently voted in favour of the EU Cross-border Healthcare Directive. This Directive, once implemented, will provide a legal basis for an improved exchange of HTA-related information between member states. This legal basis is encapsulated in Article 14 which states that:

> "The Union shall support and facilitate cooperation and the exchange of scientific information among Member States within a voluntary network connecting national authorities or bodies responsible for health technology assessment designated by the Member States" (European Parliament 2011 p.47)

Article 14 states that the cooperation aims to support EU countries with "objective, reliable, timely, transparent, comparable and transferable" evidence regarding both efficacy and potentially effective health technologies, and to improve efficiency of HTA by reducing duplication of efforts (European Parliament 2011 p. 47).

Having examined four key HTA bodies within Europe, it is relevant to consider the implications of this research for the strengthening of a European HTA network. An interesting finding is that the number of technologies appraised by all four HTA bodies for the same indication within the same year is limited to 7 out of 348 technologies, and 26 out of 348 technologies when the time window is expanded to 5 years. Thus, the actual proportion of technologies that are appraised by all four HTA bodies is limited. These results would suggest that, in addition to harmonising the assessment of specific aspects of the HTA process, there would be a need to harmonise the technologies selected. Otherwise, the usefulness of adapting one appraisal from another would be severely limited by the difference in timing of availability of such appraisals, and the risk of information being out-dated across adaptations. The analyses of HTA bodies conducted within this thesis imply that the nature of EU-wide HTA reports would need to be carefully considered and planned to contribute to national HTA initiatives, given differences in the timing and scoping considered as part of the assessment.

### 9.6 Recommendations for future research

This thesis has argued that within the four European HTA bodies examined, coverage decision-making is a complex process influenced by clinical and economic evidence, process factors and socio-economic context factors. One important area for further

research would be to investigate, in more detail, a method for standardising coverage decisions made by HTA bodies in Europe, so as to further improve the platform for evaluation of such decisions across European HTA bodies. O'Neill and Devlin (2010) proposed a methodology to standardise coverage decisions by calculating 'M' – the proportion of patients covered by public funding relative to the total eligible population. One of the key limitations recognised by O'Neill and Devlin (2010) is access to the data needed to calculate 'M'. Data collection relevant to 'M' and the application of 'M' to additional HTA bodies in Europe may represent a fruitful avenue for further research.

Furthermore, exploring to what degree, and through what mechanisms, coverage decisions by one HTA body can impact on another such body represents another potentially fruitful area for further research. Such research could be of particular relevance to improve the understanding of the factors that can help explain coverage decision-making across HTA bodies. By exploring the degree and mechanisms (whether formal or informal) through which HTA bodies influence one another other, it would be possible to provide an additional perspective on how European HTA bodies interact and the factors driving their decisions.

## 9.7 References

Achilladelis, B., and Antonakis, N. 2001. The dynamics of technological innovation: the case of the pharmaceutical industry. *Research Policy* 30, 535–588

Attridge, J .2006. Discussion Paper 2: Innovation Models in the Biopharmaceuticals Sector, part of Innovation Models and Their Application to the Pharmaceuticals Sector. http://www3.imperial.ac.uk/portal/pls/portallive/docs/1/7290711.PDF

Barbieri, M., N. Hawkins, and M. Sculpher. 2009. Who does the numbers? The role of third-party technology assessment to inform health systems' decision-making about the funding of health technologies. *Value Health* 12 (2):193-201.

Bouvenot G. 2006. Service médical rendu des médicaments. *Revue du Rhumatisme* 73:409-12.

Cazzola et al. 2010. Italian real-life experience of omalizumab. *Respiratory Medicine* 104(10):1410-1416.

Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akerhurst, R., Buxton, M., Brazier, J., and O'Hagan, T. 2005. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* 14: 339–347.

Clement, F. M., A. Harris, J. J. Li, K. Yong, K. M. Lee, and B. J. Manns. 2009. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 302 (13):1437-43.

Dakin, H.A., N.J. Devlin, and I.A.O. Odeyemi. 2006. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. *Health Policy* 77:352-367.

De Cock, J (Editor). 2010. A call to make valuable innovative medicines accessible in the European Union. Belgian Presidency of the Council of the European Union. September 2010.

Delamothe, T. 2008. Universality, equity, and quality of care. *BMJ* 336:1277-1281.

Department of Health, Medicines, Pharmacy & Industry Group. 2010. A new value-based approach to the pricing of branded medicines: a consultation. Crown copyright 2010.

Devlin, N., and D. Parkin. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 13 (5):437-52.

Eichler, H., B. Bloechl-Daum, E. Abadie, D. Barnett, F. Konig, and S. Pearson. 2010. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nature Reviews Drug Discovery* 9:277-291.

European Parliament 2011. Position of the European Parliament adopted at second reading on 19 January 2011 with a view to the adoption of Directive 2011/.../EU of the European Parliament and of the Council on the application of patients' rights in cross-border healthcare. http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2011-0007#title2

Gubb, J. 2008. Why the NHS is the sick man of Europe. *Civitas Rev* 5(1):1-11.

HAS. 2009. Rechercher un avis. http://www.has-sante.fr/portail/jcms/c_5267/actes-medicaments-dispositifs-medicaux?catName=true&replaceFileDoc=false&searchInFiles=true&portlet=c_63468&cid=c_5267&text=&dateType=pdate&dateSince=&dateSince_user=&dateSince_unit=86400000&beginDay=1&beginMonth=0&beginYear=2004&endDay=31&endMonth=5&endYear=2009&typesf=generated.AVISMedicament&opSearch=Lancer+la+recherche). Viewed 2 July 2009.

Haynes, B. 1999. Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. *BMJ* 319(7211):652-3.

Herlanda, K., J. Akselsenb, O. Henning Skjønsbergc, and L. Bjermerd 2005. How representative are clinical study patients with asthma or COPD for a larger "real life" population of patients with obstructive lung disease?, *Respir Med* 99: 11–19.

Hutton, J., C. McGrath, J.M. Frybourg, M. Tremblay, E. Bramley-Harker, and C. Henshall. 2006. Framework for describing and classifying decision-making systems using thechnology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal of Technology Assessment in Health Care* 22 (1):10-18.

Kanavos, P, E. Nicod, S. van den Aardweg, and S. Pomedli. 2010. The impact of health technology assessments: an international comparison. *Euro Observer* (4).

Koopmanschap M.A., E.A. Stolk, A.H.E. Koolman. 2010. Dear Policymaker: Have you made up your mind? A discrete choice experiment. *International Journal of Technology Assessment in Health Care* 26(2):198-204.

Kristensen F.B., K. Lampe, D.L. Chase, S.H. Lee-Robin, C. Wild , M. Moharra, M.V. Garrido , C.P. Nielsen, J.A. Røttingen, S.A. Neikter, M.L. Bistrup; European network for Health Technology Assessment (EUnetHTA). 2009. Practical tools and methods for health technology assessment in Europe: Structures, methodologies, and tools developed by the European network for Health Technology Assessment, EUnetHTA. *International Journal of Technology Assessment in Health Care*, 25: 1-8.

le Polain M, M. Franken, M. Koopmanschap, and I. Cleemput. 2010. Drug reimbursement systems: international comparison and policy recommendations. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE). 2010. KCE Reports 147C. D/2010/10.273/90

Leatherman, S., and K. Sutherland. 2008. The quest for quality: refining the NHS reforms. London: Nuffield Trust, 2008.

Lexchin, J., and B. Mintzes. 2008. Medicine reimbursement recommendations in Canada, Australia, and Scotland. *Am J Manag Care* 14 (9):581-8.

McGuire, A., and M. Litt. 2003. UK Budgetary systems and new health care technologies. *Value in Health* 6: S64-S73.

Mears A., T. Kendall, G. Strathdee, R. Sinfield and I. Aldridge. 2008. Progress on NICE guideline implementation in mental health trusts: meta-analyses. *The Psychiatrist* 32: 383-387

Morgan, S., Lopert, R., and Greyson, D. 2008. Toward a definition of pharmaceutical innovation. *Open Medicine* 2(1):e4–7

National Institute for Health and Clinical Excellence. 2008. Guide to the methods of technology appraisal Issued: June 2008. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

———. 2009. Supplementary advice to the Appraisal Committees. http://www.nice.org.uk/media/E4A/79/SupplementaryAdviceTACEoL.pdf

Neumann, P.J., M.F. Drummond, B. Jönsson, B.R. Luce, J.S. Schwartz, , U. Siebert, S. Sullivan. 2010. Are Key Principles for improved health technology assessment supported and used by health technology assessment organizations? *International Journal of Technology Assessment in Health Care* 26(1): 71-78.

O'Neill, P., and N. J. Devlin. 2010. An analysis of NICE's 'restricted' (or 'optimized') decisions. *Pharmacoeconomics* 28 (11):987-93.

OECD. 2005. *Health Technologies and Decision Making*. The OECD Health Project, OECD Publishing.

Permanand, G. 2006. *EU pharmaceutical regulation: the politics of policy-making*. Manchester: Manchester University Press

Pronk M.H., G.J. Bonsel. 2004. Out-patient drug policy by clinical assessment rather than financial constraints? The gate-keeping function of the out-patient drug reimbursement system in The Netherlands. *The European Journal of Health Economics* 5(3):274-7.

Raftery, J. 2006. Review of NICE's recommendations, 1999-2005. *BMJ* 332 (7552):1266-8.

Saltman, R.B., V. Bankauskaite, and K. Vrangbaek (Eds). 2007. *Decentralization in health care.* World Health Organization on behalf of the European Observatory on Health Systems and Policies. Maidenhead, UK: Open University Press.

SMC. 2008. Scottish Medicines Consortium. Annual Report 2008. Viewed on January 15th 2011. http://www.scottishmedicines.org.uk/files/NHS_SMC_AR_08.pdf

———. 2011. What we Do. Accessed February 20 2011, http://www.scottishmedicines.org.uk/About_SMC/What_we_do/index, Viewed on January 15th 2011.

Sermet, C. 2007. La prise en compte de l'innovation thérapeutique dans les politiques de prix et de remboursement des médicaments - Une approche internationale. *RFAS* 3-4:319-341.

Sorenson, C., M. Drummond, and P. Kanavos. 2008. Ensuring value for Money in Health Care: the role of HTA in the European Union. Cornwall: World Health Organization 2008, on behalf of the European Observatory on Health Systems and Policies.

Stolk E.A., M.J. Poley. 2005. Criteria for determining a basic health services package. Recent developments in The Netherlands. *The European journal of health economics* 6(1):2-7.

Vegter, S., M.H. Rozenbaum, R.Postema, K. Tolley, and M.J. Postma. 2010. Review of Regulatory Recommendations for Orphan Drug Submissions in the Netherlands and Scotland: Focus on the Underlying Pharmacoeconomic Evaluations. *Clinical Therapeutics* 32(9):1651-1661.

Wilking, N, and B Jonsson. 2005. A pan-European comparison regarding patient access to cancer drugs. Stockholm: Karolinksa Institutet and Stockholm School of Economics.

Weiss, C. 1979. The Many Meanings of Research Utilization. *Public Administration review* 39 (5):426-431.

# Appendices

# Contents

# A. Chapter 4 Appendices

## NICE coverage decisions 2004-2009: List of Technology Appraisals included for analysis

| Technology Appraised | Ref | Year Guidance was issued |
|---|---|---|
| Pegylated interferon (in ribavirin combination therapy) or as monotherapy | TA75 | 2004 |
| gabapentin | TA76 | 2004 |
| zaleplon | TA77 | 2004 |
| clopidogrel | TA80 | 2004 |
| pimecrolimus | TA82 | 2004 |
| Drotrecogin alfa | TA84 | 2004 |
| Basiliximab | TA85 | 2004 |
| imatinib | TA86 | 2004 |
| clopidogrel | TA90 | 2005 |
| paclitaxel | TA91 | 2005 |
| irinotecan | TA93 | 2005 |
| atorvastatin | TA94 | 2006 |
| adefovir dipivoxil | TA96 | 2006 |
| capecitabine | TA100 | 2006 |
| docetaxel | TA101 | 2006 |
| efalizumab | TA103 | 2006 |
| etanercept | TA104 | 2006 |
| peginterferon alfa-2a | TA106 | 2006 |
| trastuzumab | TA107 | 2006 |
| paclitaxel | TA108 | 2006 |
| docetaxel | TA109 | 2006 |
| rituximab | TA110 | 2006 |
| donepezil | TA111 | 2007 |
| anastrozole | TA112 | 2006 |
| buprenorphine | TA114 | 2007 |
| naltrexone | TA115 | 2007 |
| gemcitabine | TA116 | 2007 |
| cinacalcet | TA117 | 2007 |
| bevacizumab | TA118 | 2007 |
| fludarabine | TA119 | 2007 |
| alteplase | TA122 | 2007 |
| varenicline | TA123 | 2007 |
| pemetrexed | TA124 | 2007 |
| adalimumab | TA125 | 2007 |
| rituximab | TA126 | 2007 |
| natalizumab | TA127 | 2007 |
| bortezomib | TA129 | 2007 |
| adalimumab | TA130 | 2007 |
| ezetimibe | TA132 | 2007 |
| omalizumab | TA133 | 2007 |
| infliximab | TA134 | 2008 |
| pemetrexed | TA135 | 2008 |
| rituximab | TA137 | 2008 |
| infliximab | TA140 | 2008 |
| abatacept | TA141 | 2008 |
| Darbepoetin alfa | TA142 | 2008 |
| adalimumab | TA143 | 2008 |
| cetuximab | TA145 | 2008 |
| adalimumab | TA146 | 2008 |
| entecavir | TA153 | 2008 |
| telbivudine | TA154 | 2008 |
| pegaptanib | TA155 | 2008 |
| dabigatran | TA157 | 2008 |

| | | |
|---|---|---|
| amantandine | TA158 | 2008 |
| alendronate | TA160 | 2008 |
| alendronate | TA161 | 2008 |
| erlotinib | TA162 | 2008 |
| infliximab | TA163 | 2008 |
| febuxostat | TA164 | 2008 |
| amantandine | TA168 | 2009 |
| sunitinib | TA169 | 2009 |
| rivaroxaban | TA170 | 2009 |
| lenalidomide | TA171 | 2009 |
| cetuximab | TA172 | 2009 |

**NICE Data-set: Missing Data**

Between January 2004-June 2009, the National Institute for Health and Clinical Excellence (NICE) made 118 funding decisions – either recommending, restricting or not recommending use of NHS resources to fund new health technologies (in this research analysis, the health technologies are restricted to pharmaceutical products).

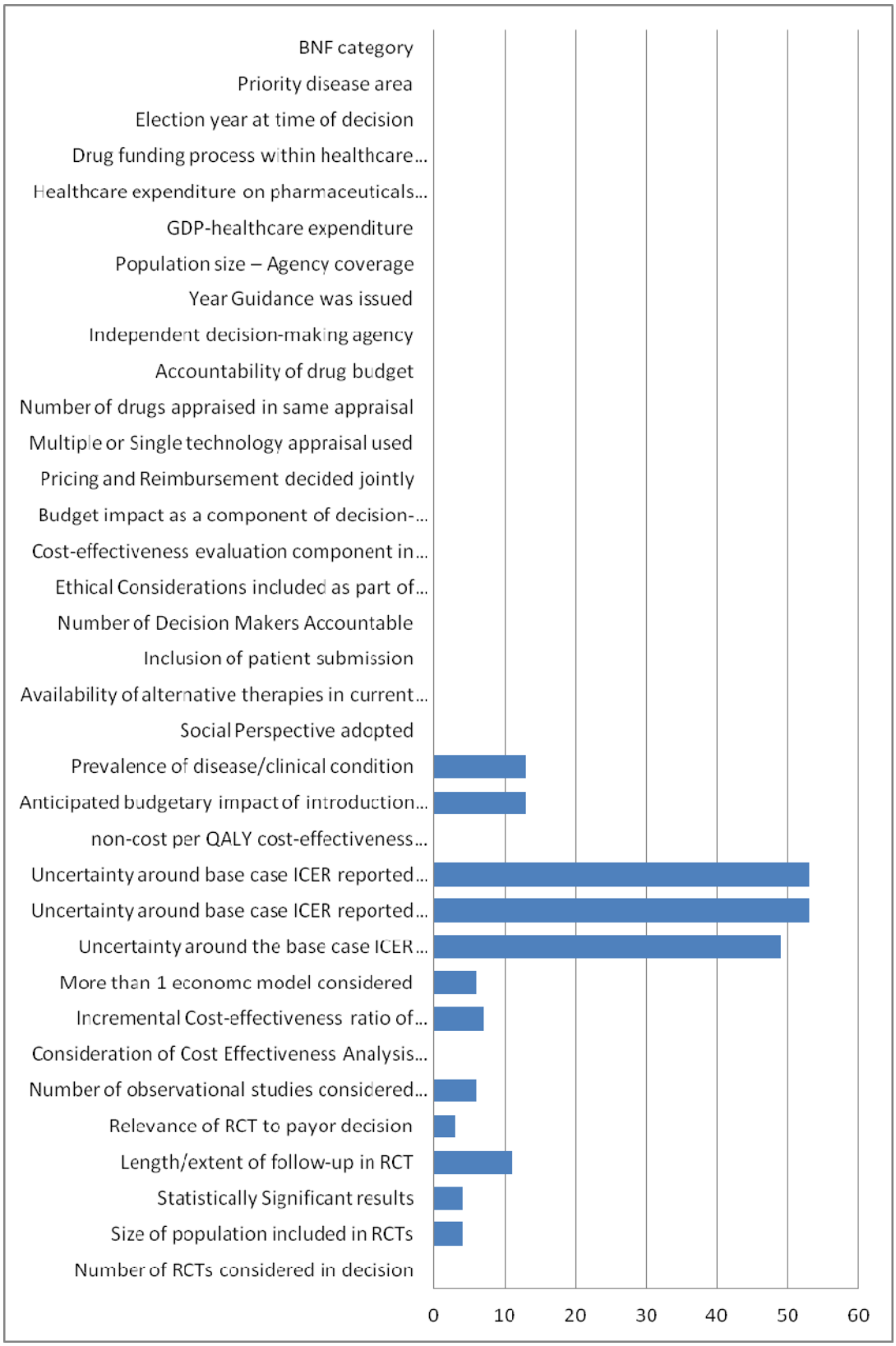
A data set of information pertaining to NICE appraisals was created collecting information on variables relating to (i) the clinical and economic characteristics of the technology under appraisal, as well as information on (ii) the process used to come to a decision, and (iii) the socio-economic context in which these decisions were made.

In order to prepare the data set for analysis, understanding the data set is important. Here, the aim is to characterise the presence of missing data across variables and decisions to inform the need for imputing missing data.

*Distribution of Missing Data within NICE data set*

In the total NICE sample, there is about 8% of entries were incomplete. The distribution of missing data **across each variable** was examined. The total number of observations per variable is 118. The variables with the highest number of 'not reported' information are those related to the economic characteristics of the technology. 11 variables have no missing data. The extent of missing data was also examined **across appraisals**. These appraisals span almost 5 years of decision making. The range of variables where information was missing ranged from 0-9. The mean number of variables where information was missing was 3.

**Figure A.1 NICE Distribution of missing data by variable (n=118)**



**NICE Dataset – Testing of Normality Assumption**

*Method*

To determine the relevant statistical tests to use in assessing the significance of differences observed between means, it was necessary to assess whether the normality assumption was valid for the variables under consideration. This would then determine the use of parametric or non-parametric tests. For all variables, the sample is >30. It has been suggested that when analysing sample sizes of >30, even when the normality assumption is violated, parametric tests may still be performed (Pallant 2007, SPSS Survival Manual, 3rd edn, Maidenhead, OUP/McGraw-Hill). Prior to making a decision on which variables to apply parametric or non-parametric tests, the distribution for each variable will be further examined.

To test the normality assumption the following was performed for each non-categorical variable:

- Skeweness was calculated
- Kurtosis was calculated
- The standardized normal probability plot (P-P plot) was graphed

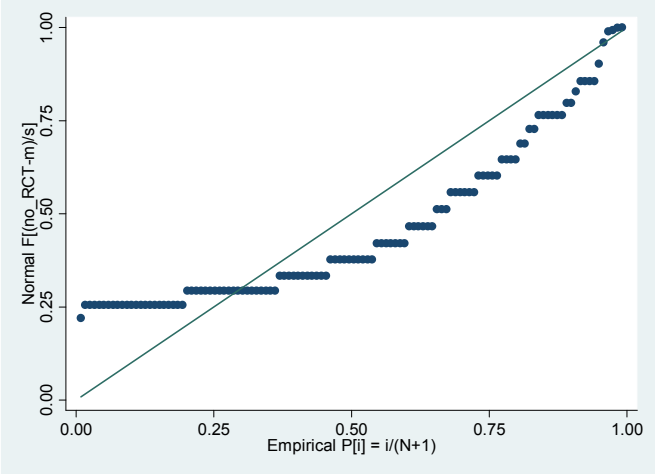
The graph and calculations are presented for each non-categorical variable and an assessment is made of whether the normality assumption is met or not. A variable will be considered to meet the normal distribution assumption if

- skewness is within the range ±1
- Kurtosis value is within range ±3
- P-P plot shows distribution of dataset is approximately linear

In those circumstances where there is inconsistency between the three tests, if 2 of the tests suggest normal distribution, then it will be considered as such (but then tested using both parametric and non-parametric tests in sensitivity analyses).

**Number of RCTs considered in appraisal (No_RCT)**

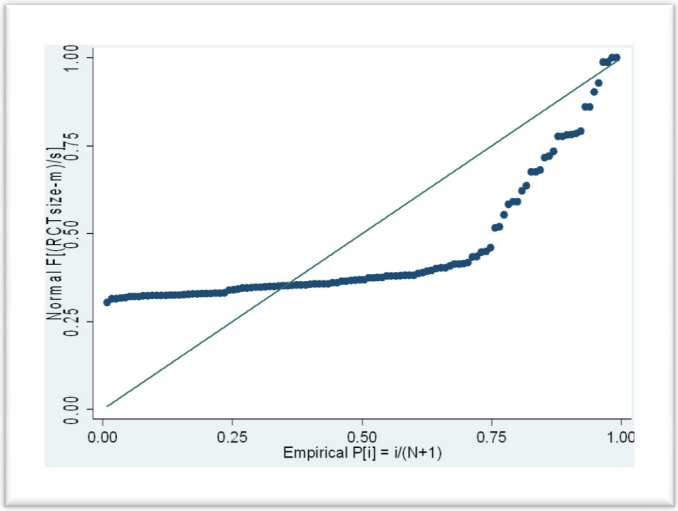
- Skewness 4.402407
- Kurtosis 30.86422




This variable is **not** normally distributed.

**Mean sample size of RCTs considered in appraisal (RCTsize)**

- Skewness 4.995812
- Kurtosis 33.70005



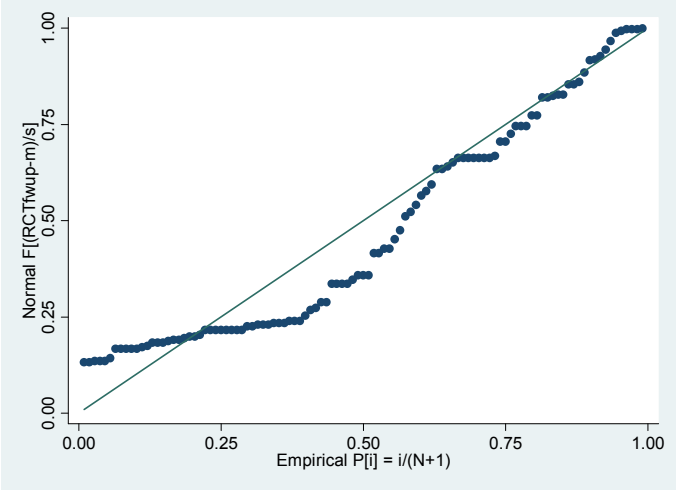

This variable is **not** normally distributed.

**Superiority demonstrated**

Normality test not performed as this is a categorical variable.

**Duration of RCT (RCTfwup)**

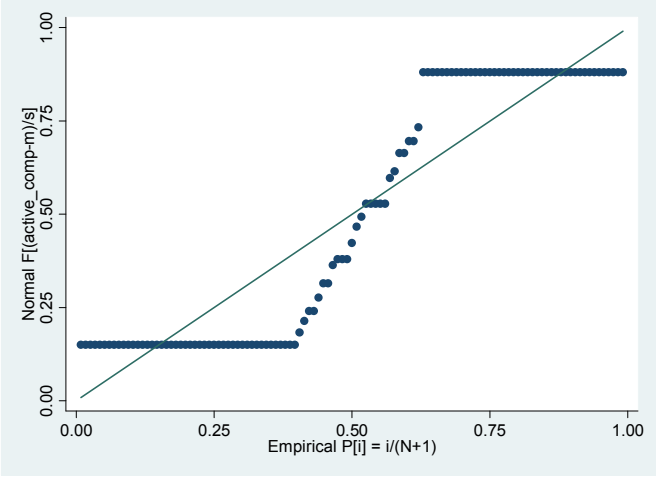
- Skewness 1.151972
- Kurtosis 3.870274




This variable is **not** normally distributed.

**Active-Comp**

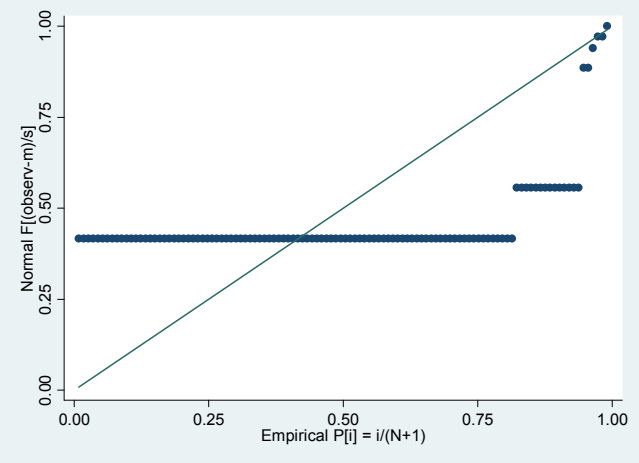
- Skewness .1479741
- Kurtosis 1.216005




This variable **is** normally distributed.

**Observational Studies**

- Skewness 8.428139
- Kurtosis 80.60793



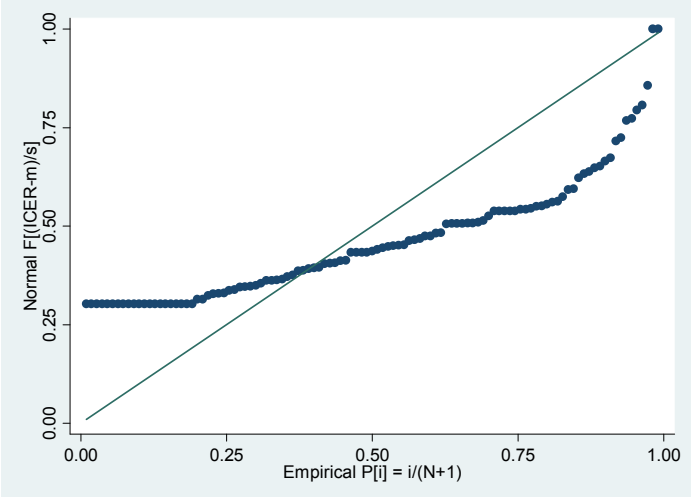

This variable is **not** normally distributed.

**CUA performed**

Normality test not performed as this is a categorical variable.

**ICER**

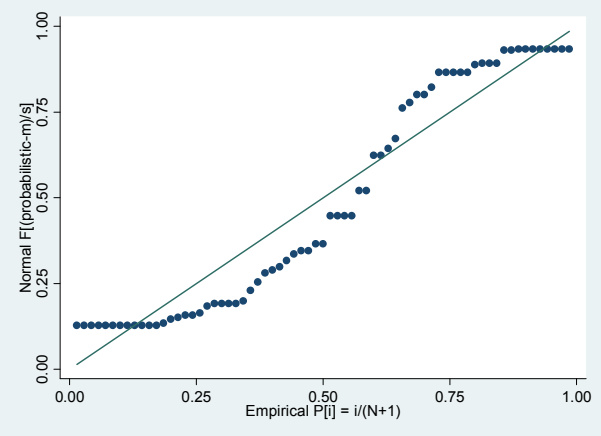
- Skewness 6.973113
- Kurtosis 59.67012




This variable is **not** normally distributed.

**ICER SA – Probabilistic**

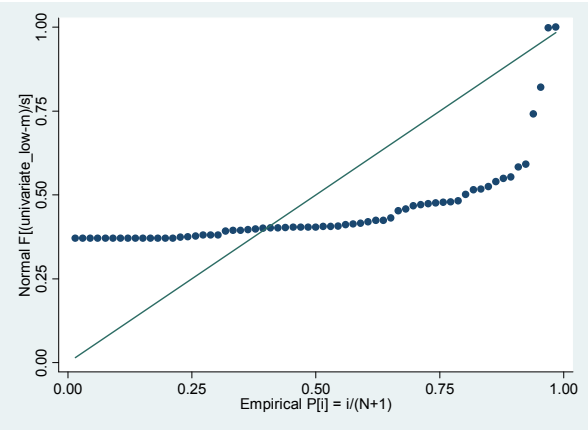
- Skewness .3108645
- Kurtosis 1.50603




This variable is normally distributed.

**ICER SA – Univariate Low**

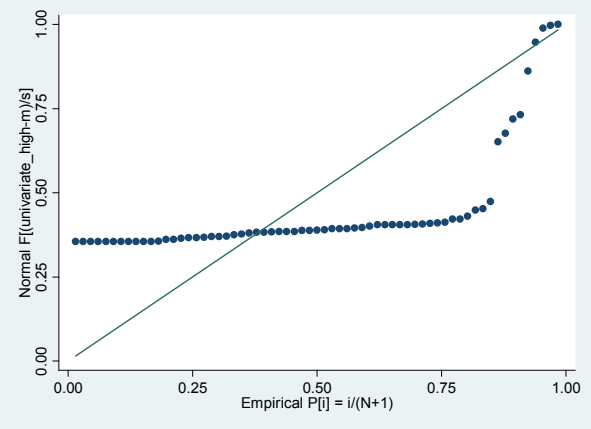
- Skewness 6.018749
- Kurtosis 41.22995




This variable is **not** normally distributed.

**ICER SA – Univariate high**

- Skewness 4.760234
- Kurtosis 28.4462

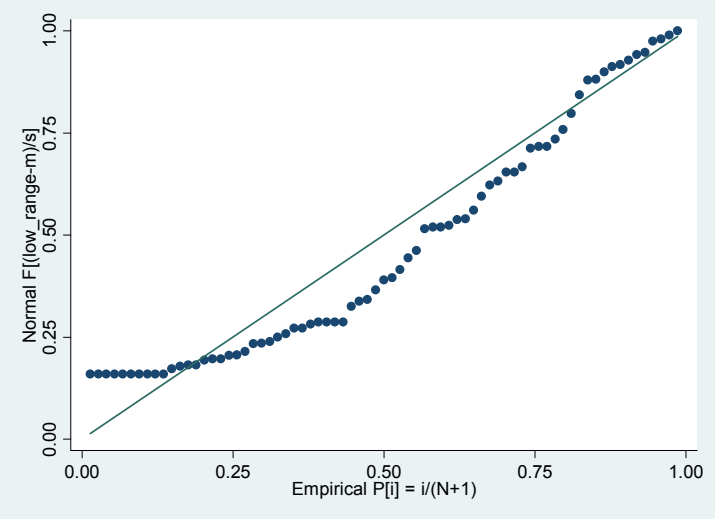



**Multiple Models submitted**

Normality test not performed as this is a categorical variable.

**Multiple Models: Range of ICERs - LOW**

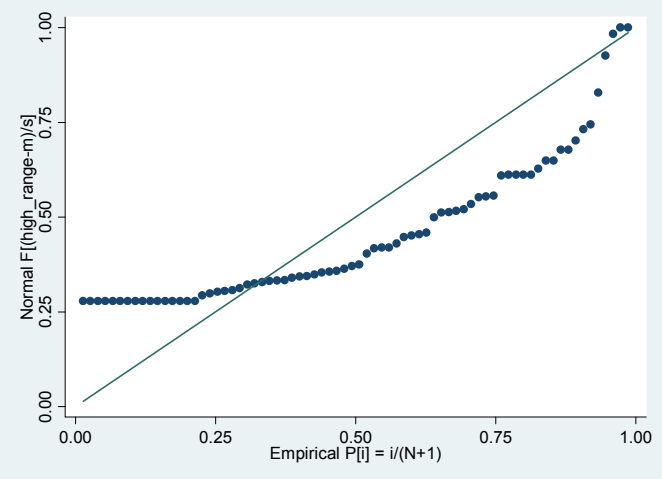
- Skewness 1.325258
- Kurtosis 4.990431




This variable **could** be considered to approximate to a normal distribution.

**Multiple Models: Range of ICERs – HIGH**

- Skewness 4.245608
- Kurtosis 24.62567



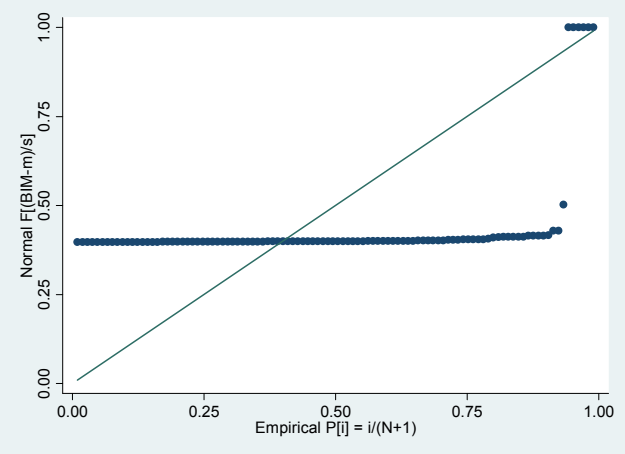

This variable is **not** normally distributed.

**Non_CUA submitted**

Normality test not performed as this is a categorical variable.

**BIM**

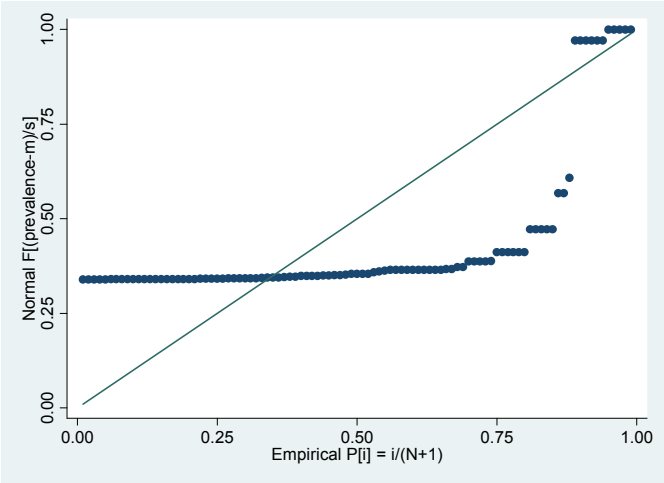
- Skewness 3.788561
- Kurtosis 15.36788




This variable is **not** normally distributed.

**Prevalence**

- Skewness 2.785481
- Kurtosis 9.566752




This variable is **not** normally distributed.

**Societal**

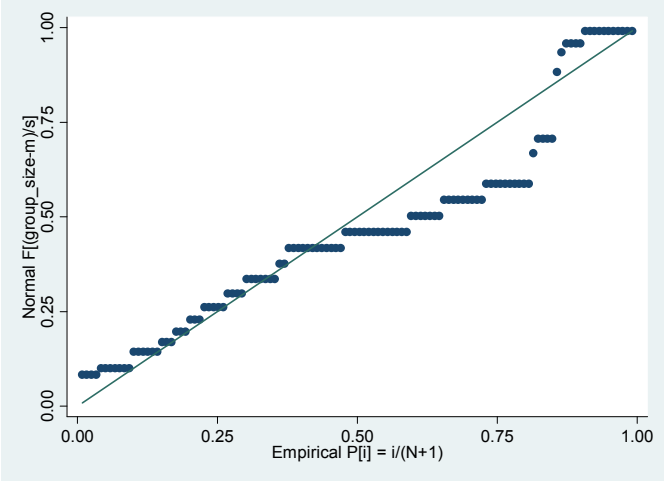
Normality test not performed as this is a categorical variable.

**Alternative Available**

Normality test not performed as this is a categorical variable.

**Group_Size**

- Skewness 1.122694
- Kurtosis 3.750685




This variable **could** be considered to approximate to a normal distribution.

**Ethics**

Normality test not performed as this is a categorical variable.

**Cost-Effectiveness part of Process**

Normality test not performed as this is a categorical variable.

**BIM part of Process**

Normality test not performed as this is a categorical variable.

**Joint reimbursement & pricing decision**

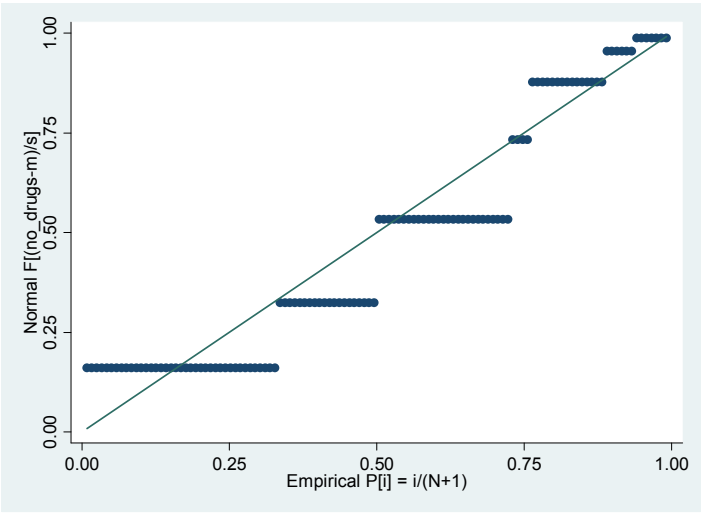
Normality test not performed as this is a categorical variable.

**MTA vs STA**

Normality test not performed as this is a categorical variable.

**No of Drugs appraised**

- Skewness .7873726
- Kurtosis 2.516377




Perhaps this is actually a categorical variable?

**Accountability for drug budget**

Normality test not performed as this is a categorical variable.

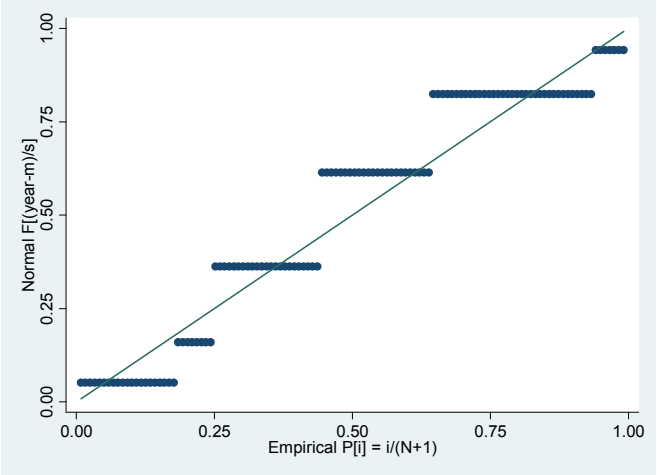
**Independence of HTA agency from MoH**

Normality test not performed as this is a categorical variable.

**Year of appraisal**

Skewness -.3891203

Kurtosis 1.978506

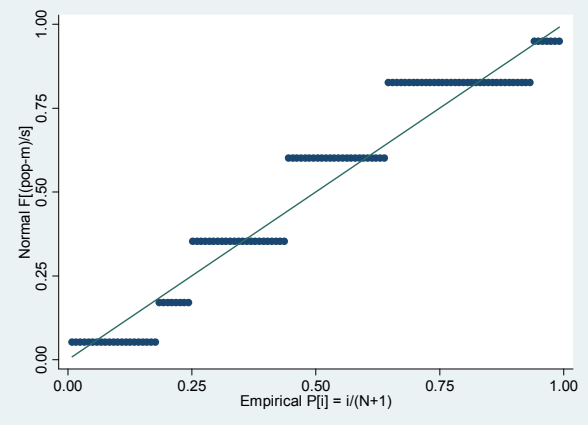



Perhaps this is actually a categorical variable?

**Population under HTA remit**

Skewness -.327666

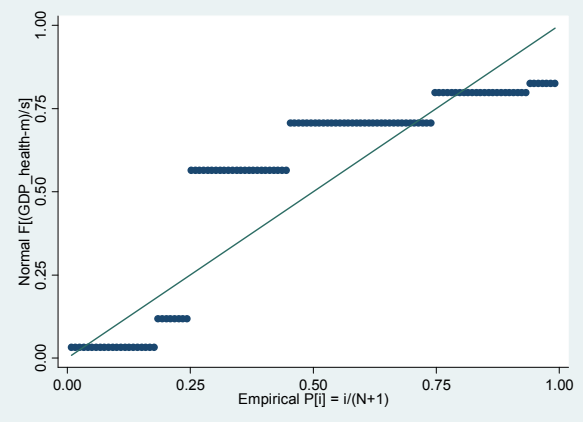
Kurtosis 1.985149




Perhaps this is actually a categorical variable?

**% of GDP expenditure on health**

- Skewness -1.063346
- Kurtosis 2.47445

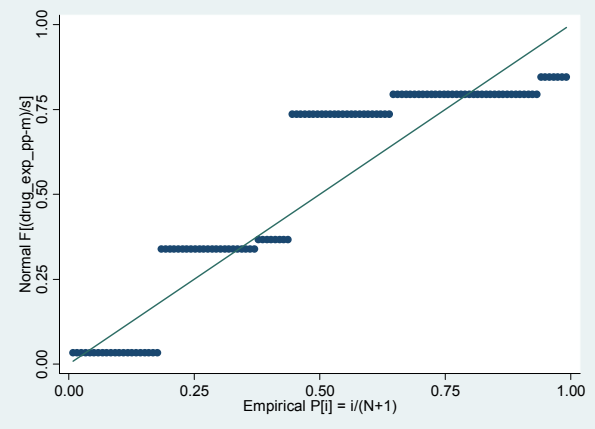



Perhaps this is actually a categorical variable?

**Drug expenditure per person**

Skewness -.8687147

Kurtosis 2.346379




**Funding mechanism – centralised or decentralised**

Normality test not performed as this is a categorical variable.

**Election year**

Normality test not performed as this is a categorical variable.

**Priority Disease Area**

Normality test not performed as this is a categorical variable.

**BNF Category**

Normality test not performed as this is a categorical variable.

**NICE dataset: descriptive statistics**

| | Chi2 | ANOVA | T-Test (Rec vs Res) | T-Test (Res vs NR) | Kruskal-Wallis |
|---|---|---|---|---|---|
| Number of RCTs considered in decision | | 0.2152 | 0.7448 | 0.0002 | 0.021 |
| Size of population included in RCTs | | 0.0221 | 0.2017 | 0.8035 | 0.0504 |
| Statistically Significant results - yes | 0.013 | | | | |
| no | 0.143 | | | | |
| inconsistent | 0.207 | | | | |
| Length/extent of follow-up in RCT | | 0.3069 | 0.0411 | 0.3734 | 0.2129 |
| Relevance of RCT to payor decision | | 0.7077 | 0.015 | 0.7454 | 0.0544 |
| Number of observational studies considered in guidance | | 0.2569 | 0.2247 | 0.0032 | 0.1361 |
| Consideration of Cost Utility Analysis in guidance | 0.187 | | | | |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | | - | 0.054 | 0.1158 | 0.0001 |
| More than one CUA submitted | 0.716 | | | | |
| If More than one CUA submitted - low range | | 0.0024 | 0.0448 | 0.1697 | 0.0229 |
| If More than one CUA submitted - high range | | - | 0.3745 | 0.1332 | 0.0002 |
| Uncertainty around base case ICER reported in submission (univariate) Low | | 0.033 | 0.044 | 0.000 | 0.001 |
| Uncertainty around base case ICER reported in submission (univariate) High | | - | 0.1271 | 0.2866 | 0.0496 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | | - | 0.3215 | 0.0893 | 0.0002 |
| Non-CUA analyses submitted | 0.056 | | | | |
| Anticipated budgetary impact of introduction of new technology in health care system | | 0.0772 | 0.046 | 0.4718 | 0.4057 |
| Prevalence of disease/clinical condition | | 0.0027 | 0.0026 | 0.4413 | 0.183 |
| Societal Perspective adopted | 0.77 | | | | |
| Availability of alternative therapies in current treatment setting. | 0.935 | | | | |
| Inclusion of patient submission | 0.431 | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Number of Decision Makers Accountable | | 0.5187 | 0.1522 | 0.2735 | 0.4027 |
| Cost-effectiveness evaluation component in process | - | - | - | - | - |
| Budget impact as a component of decision-making process | - | - | - | - | - |
| Price of technology known during appraisal | - | - | - | - | - |
| Use of STA process | 0.031 | | | | |
| Number of drugs appraised in same appraisal | | 0.7214 | 0.0001 | 0.0015 | 0.0006 |
| Accountability of drug budget | - | - | - | - | - |
| Independence of decision-making agency | - | - | - | - | - |
| Date guidance was issued | | 0.0166 | 0.4447 | 0.0714 | 0.0501 |
| Population size – Agency coverage | | 0.0166 | 0.4191 | 0.0714 | 0.0501 |
| GDP-healthcare expenditure | | 0.0166 | 0.2573 | 0.1702 | 0.2022 |
| Healthcare expenditure on pharmaceuticals | | 0.0166 | 0.4573 | 0.0605 | 0.0247 |
| Election year at time of decision | 0.971 | | | | |
| Priority disease area | 0.702 | | | | |
| Orphan Designated | 0.77 | | | | |
| cardiovascular system | 0.322 | | | | |
| central nervous system | 0.067 | | | | |
| ear, nose and oropharynx | n/a | | | | |
| endocrine system | 0.699 | | | | |
| eye | 0.185 | | | | |
| gastro-intestinal system | 0.306 | | | | |
| infections | 0.183 | | | | |
| malignant disease and immunosuppression | 0.122 | | | | |
| musculoskeletal and joint diseases | 0.754 | | | | |
| nutrition and blood | 0.335 | | | | |
| obstetrics, gynaecology, and urinary-tract disorders | n/a | | | | |
| respiratory system | 0.699 | | | | |
| skin | 0.84 | | | | |

## NICE dataset: preliminary multivariate model

The pseudo R-squared for this model was 0.5528, which suggests that it explains 55% of the variability observed in NICE coverage decisions.

**Multivariate analysis of NICE coverage decisions 2004-2009: preliminary model (n=118)**

| Restricted | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | 3.059583 | 0.207 | -1.694669 | 7.813835 |
| Lack of clinical superiority in RCT | 4.830348 | 0.065 | -0.3064264 | 9.967122 |
| No clinical superiority demonstrated in RCT | 4.781009 | 0.044 | 1.29E-01 | 9.432898 |
| Number of RCTs | -0.0642417 | 0.183 | -0.1587155 | 0.030232 |
| RCT duration of follow-up | -0.0002126 | 0.976 | -0.0140222 | 0.013597 |
| Use of active comparator in RCT | -0.9846096 | 0.270 | -2.732692 | 0.7634723 |
| ICER | 0.0000335 | 0.246 | -0.0000231 | 0.0000901 |
| Uncertainty range around the ICER | -0.00000423 | 0.083 | -9.02E-06 | 0.000000555 |
| Probability of ICER < 30,000£ | -1.630215 | 0.328 | -4.897945 | 1.637514 |
| Use of non cost-utility analysis | 0.9884623 | 0.350 | -1.083165 | 3.06009 |
| Budget impact of new technology | -0.0000693 | 0.902 | -0.0011751 | 0.0010365 |
| Disease prevalence | 0.000000153 | 0.386 | -0.000000193 | 0.000000498 |
| Use of MTA process vs. STA process | -0.4195647 | 0.689 | -2.47E+00 | 1.633965 |
| Number of technologies appraised simultaneously | 0.25413 | 0.355 | -0.2848913 | 0.7931513 |
| Year of Appraisal | 0.8785522 | 0.033 | 0.0703909 | 1.686714 |
| Technology indicated for cancer treatment | -0.491717 | 0.644 | -2.578527 | 1.595093 |
| No information on ICER | 23.35165 | 0.000 | 17.92525 | 28.77805 |
| No information on uncertainty around ICER | -0.1537224 | 0.852 | -1.76E+00 | 1.456904 |
| No information on probabilistic sensitivity analysis around ICER | 1.639982 | 0.049 | 0.0040381 | 3.275925 |
| Constant | -1766.018 | 0.033 | -3386.276 | -145.7601 |
| Not Recommended | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -6.392912 | 0.153 | -15.15638 | 2.37056 |
| Lack of clinical superiority in RCT | -3.462322 | 0.459 | -12.61755 | 5.692911 |
| No clinical superiority demonstrated in RCT | 0.4864004 | 0.859 | -4.89E+00 | 5.866077 |
| Number of RCTs | -0.156669 | 0.460 | -0.5726012 | 0.2592633 |
| RCT duration of follow-up | 0.0456999 | 0.082 | -0.0058245 | 0.0972242 |
| Use of active comparator in RCT | -0.883584 | 0.643 | -4.618792 | 2.851624 |
| ICER | 0.0002064 | 0.044 | 0.00000565 | 0.0004072 |
| Uncertainty range around the ICER | 0.00000244 | 0.268 | -1.88E-06 | 0.00000677 |
| Probability of ICER < 30,000£ | -1.847847 | 0.555 | -7.990323 | 4.29463 |
| Use of non cost-utility analysis | -56.91792 | . | . | . |
| Budget impact of new technology | 0.0000328 | 0.962 | -0.0013099 | 0.0013755 |
| Disease prevalence | 0.00000032 | 0.362 | -0.000000368 | 0.00000101 |
| Use of MTA process versus STA process | 0.4392 | 0.833 | -3.64E+00 | 4.521612 |
| Number of technologies appraised simultaneously | -1.471622 | 0.096 | -3.203731 | 0.2604862 |
| Year of Appraisal | 0.9634825 | 0.174 | -0.4260984 | 2.353063 |
| Technology indicated for cancer treatment | -3.519016 | 0.165 | -8.489831 | 1.451799 |
| No information on ICER | 28.69518 | . | . | . |
| No information on uncertainty around ICER | 1.780973 | 0.511 | -3.532484 | 7.094431 |

| | | | | |
|---|---|---|---|---|
| No information on probabilistic sensitivity analysis around ICER | 2.496638 | 0.173 | -1.09E+00 | 6.083571 |
| Constant | -1939.318 | 0.173 | -4728.422 | 849.7874 |

Note: Recommended technologies are the reference case

**Sensitivity Analysis**

In a third sensitivity analysis, a categorical ICER variable was used as opposed to a continuous ICER variable used in the base case analysis. This binary variable was created by distinguishing between those ICERs above £30,000 and those technologies with ICERs below or equal to £30,000. This model yielded a similar pseudo R-squared to the base case model (0.2867 for the sensitivity analysis model and 0.2642 for the base case) (Table 4.8). In this analysis, being cost-effective (i.e. with ICER of less than £30,000), statistically significantly decreased the log odds of a restriction or a non recommendation. When technologies were associated with ICERs that were above £30,000, this decreased the odds of a recommendation relative to a restriction (p=0.002) and relative to a non-recommendation (p<0.0001). The direction of the effect and statistical significance of the remaining three variables was similar between this sensitivity analysis and the base case analysis. This suggests that inclusion of the dominant/dominated technologies is important for fully capturing the impact of cost-effectiveness on NICE decision-making.

**Table 4.1 Sensitivity Analysis 3. Multivariate analysis of NICE coverage decisions 2004-2009: alternative model using a categorical ICER variable (n=118)**

| Restricted | Log Odds | P value | 95% Confindence Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -1.611471 | 0.006 | -2.766507 | -0.4564342 |
| ICER below 30,000 | -2.064262 | 0.002 | -3.39123 | -0.7372939 |
| Number of technologies appraised simultaneously | 0.6299023 | 0.001 | 2.62E-01 | 0.9977285 |
| Year of Appraisal | 0.5572081 | 0.010 | 0.1339199 | 0.9804963 |
| Constant | -1116.647 | 0.010 | -1965.812 | -267.482 |
| Not Recommended | Log Odds | P value | 95% Confidence Interval | |
| Clinical superiority demonstrated in RCT | -2.285657 | 0.009 | -3.993731 | -0.577583 |
| ICER below 30,000 | -4.440024 | 0.000 | -6.392607 | -2.487441 |
| Number of technologies appraised simultaneously | 0.3353649 | 0.231 | -2.13E-01 | 0.8836412 |
| Year of Appraisal | 1.026511 | 0.001 | 0.4016877 | 1.651335 |
| Constant | -2058.042 | 0.001 | -3311.775 | -804.3086 |

Note: Recommended technologies are the reference case

# B. Chapter 5 Appendices

## SMC coverage decisions 2004-2009: List of Technology Appraisals included for analysis

| Technology Appraised | SMC ID | Year of appraisal |
|---|---|---|
| adalimumab | 218/05 | 2005 |
| anagrelide | 163/05 | 2005 |
| Candesartan cilexetil | 161/05 | 2005 |
| capecitabine | 193/05 | 2005 |
| Carmustine | 215/05 | 2005 |
| docetaxel | 201/05 | 2005 |
| Eplerenone | 136/04 | 2005 |
| Fosamprenavir | 188/05 | 2005 |
| Mycophenolate | 144/04 | 2005 |
| Oxaliplatin | 211/05 | 2005 |
| Palonosetron | 208/05 | 2005 |
| Pegylated interferon alfa 2a | 186/05 | 2005 |
| TachoSil | 168/05 | 2005 |
| adefovir | 54/03 | 2005 |
| anastrozole | 198/05 | 2005 |
| atomoxetine | 153/05 | 2005 |
| bivalirudin | 156/05 | 2005 |
| calcipotriol | 09/02 | 2005 |
| caspofungin | 147/04 | 2005 |
| Ciclesonide | 184/05 | 2005 |
| duloxetine | 195/05 | 2005 |
| Eflornithine | 159/05 | 2005 |
| Etanercept | 212/05 | 2005 |
| Exemestane | 210/05 | 2005 |
| Iloprost | 219/05 | 2005 |
| imiquimod | 167/05 | 2005 |
| infliximab | 101/04 | 2005 |
| letrozole | 152/05 | 2005 |
| Montelukast | 185/05 | 2005 |
| Oxybutynin | 190/05 | 2005 |
| Oxycodone | 197/05 | 2005 |
| pemetrexed | 192/05 | 2005 |
| pioglitazone | 115/04 | 2005 |
| pregabalin | 145/04 | 2005 |
| Strontium | 178/05 | 2005 |
| Valsartan | 162/05 | 2005 |
| Vinorelbine | 179/05 | 2005 |
| Voriconazole | 194/05 | 2005 |
| Zonisamide | 216/05 | 2005 |
| anagrelide | 163/05 | 2005 |
| bemiparin | 204/05 | 2005 |
| bemiparin | 205/05 | 2005 |
| bemiparin | 206/05 | 2005 |
| bemiparin | 203/05 | 2005 |
| cetuximab | 155/05 | 2005 |
| Cilostazol | 86/04 | 2005 |
| Cinacalcet | 169/05 | 2005 |
| diclofenac | 199/05 | 2005 |
| docetaxel | 209/05 | 2005 |
| Eflornithine | 159/05 | 2005 |
| Erlotinib | 220/05 | 2005 |

| | | |
|---|---|---|
| Gemcitabine/paclitaxel | 154/05 | 2005 |
| glyceryl trinitrate | 200/05 | 2005 |
| Ibritumomab | 171/05 | 2005 |
| liposomal cytarabine | 164/05 | 2005 |
| Metformin | 148/04 | 2005 |
| Modafinil | 183/05 | 2005 |
| Modafinil | 63/03 | 2005 |
| Nicotonic acid | 93/04 | 2005 |
| Pegvisomant | 158/05 | 2005 |
| pregabalin | 157/05 | 2005 |
| pregabalin | 157/05 | 2005 |
| Solifenacin | 129/04 | 2005 |
| Emtricitabine | 105/04 | 2006 |
| entecavir | 320/06 | 2006 |
| escitalopram | 253/06 | 2006 |
| gemcitabine | 154/05 | 2006 |
| Ibandronic acid | 228/05 | 2006 |
| Olopatadine | 59/03 | 2006 |
| posaconazole | 256/06 | 2006 |
| pramipexole | 247/06 | 2006 |
| zoledronic acid | 317/06 | 2006 |
| adalimumab | 300/06 | 2006 |
| anastrozole | 322/06 | 2006 |
| carglumic | 299/06 | 2006 |
| cetuximab | 279/06 | 2006 |
| daptomycin | 248/06 | 2006 |
| duloxetine | 285/06 | 2006 |
| Erlotinib | 220/05 | 2006 |
| fludarabine | 176/05 | 2006 |
| Ibandronic acid | 301/06 | 2006 |
| insulin glulisine | 298/06 | 2006 |
| letrozole | 251/06 | 2006 |
| Pegaptanib | 290/06 | 2006 |
| rituximab | 323/06 | 2006 |
| rituximab | 330/06 | 2006 |
| ropinirole | 165/05 | 2006 |
| Sildenafil | 235/06 | 2006 |
| temozolomide | 244/06 | 2006 |
| tigecycline | 276/06 | 2006 |
| tigecycline | 277/06 | 2006 |
| tipranavir | 226/06 | 2006 |
| topiramate | 297/06 | 2006 |
| Trastuzumab | 278/06 | 2006 |
| aprepitant | 242/06 | 2006 |
| bevacizumab | 221/05 | 2006 |
| bevacizumab | 221/05 | 2006 |
| bortezomib | 302/06 | 2006 |
| buprenorphine | 234/06 | 2006 |
| Cinacalcet | 169/05 | 2006 |
| co-careldopa | 316/06 | 2006 |
| esomeprazole | 257/06 | 2006 |
| esomeprazole | 274/06 | 2006 |
| Estradiol | 230/05 | 2006 |
| estradiol/drospiren | 227/05 | 2006 |
| fondaparinux | 287/06 | 2006 |
| glyceryl trinitrate | 200/05 | 2006 |
| Metformin | 148/04 | 2006 |
| mitotane | 328/06 | 2006 |
| Modafinil | 63/03 | 2006 |
| natalizumab | 329/06 | 2006 |

| | | |
|---|---|---|
| nebivolol | 214/05 | 2006 |
| Nicotonic acid | 93/04 | 2006 |
| omalizumab | 259/06 | 2006 |
| paricalcitol | 288/06 | 2006 |
| Pegvisomant | 158/05 | 2006 |
| rasagiline | 255/06 | 2006 |
| rasagiline | 255/06 | 2006 |
| rasagiline | 243/06 | 2006 |
| rasagiline | 243/06 | 2006 |
| rotigotine | 289/06 | 2006 |
| sodium oxybate | 246/06 | 2006 |
| sorafenib | 321/06 | 2006 |
| sunitinib | 275/06 | 2006 |
| temozolomide | 244/06 | 2006 |
| tipranavir | 226/05 | 2006 |
| Tramadol | 236/06 | 2006 |
| azelaic acid | 359/07 | 2007 |
| budesonide/formoterol | 362/07 | 2007 |
| busulfan | 337/06 | 2007 |
| capecitabine | 401/07 | 2007 |
| darunavir | 378/07 | 2007 |
| esomeprazole | 422/07 | 2007 |
| fondaparinux | 420/07 | 2007 |
| nebivolol | 214/05 | 2007 |
| pioglitazone | 399/07 | 2007 |
| ranibizumab | 381/07 | 2007 |
| rotigotine | 289/06 | 2007 |
| TachoSil | 344/07 | 2007 |
| varenicline | 336/06 | 2007 |
| buprenorphine/naloxone | 355/07 | 2007 |
| clopidogrel | 390/07 | 2007 |
| darifenacin | 377/07 | 2007 |
| dasatinib | 370/07 | 2007 |
| deferasirox | 347/07 | 2007 |
| dibotermin alfa | 365/07 | 2007 |
| docetaxel | 369/07 | 2007 |
| ertapenem | 404/07 | 2007 |
| ertapenem | 335/06 | 2007 |
| exenatide | 376/07 | 2007 |
| infliximab | 318/06 | 2007 |
| ivabradine | 319/06 | 2007 |
| lanthanum | 286/06 | 2007 |
| natalizumab | 329/06 | 2007 |
| omalizumab | 259/06 | 2007 |
| parathyroid | 356/07 | 2007 |
| pioglitazone | 354/07 | 2007 |
| posaconazole | 379/07 | 2007 |
| rotigotine | 392/07 | 2007 |
| sitagliptin | 408/07 | 2007 |
| sitaxentan sodium | 360/07 | 2007 |
| tacrolimus | 346/07 | 2007 |
| topotecan | 421/07 | 2007 |
| abatacept | 400/07 | 2007 |
| adalimumab | 417/07 | 2007 |
| alglucosidase alfa | 352/07 | 2007 |
| beclometasone | 166/05 | 2007 |
| betaine anhydrous | 407/07 | 2007 |
| bortezomib | 302/06 | 2007 |
| buprenorphine | 234/06 | 2007 |
| clostridium botulinum | 353/07 | 2007 |

| | | |
|---|---|---|
| dasatinib | 371/07 | 2007 |
| dexrazoxane | 361/07 | 2007 |
| erdosteine | 415/07 | 2007 |
| escitalopram | 406/07 | 2007 |
| glyceryl trinitrate | 200/05 | 2007 |
| Ibritumomab | 171/05 | 2007 |
| idursulfase | 391/07 | 2007 |
| imiquimod | 385/07 | 2007 |
| infliximab | 363/07 | 2007 |
| infliximab | 364/07 | 2007 |
| Interferon beta-1b | 345/07 | 2007 |
| levetiracetam | 395/07 | 2007 |
| levetiracetam | 396/07 | 2007 |
| levetiracetam | 397/07 | 2007 |
| lidocaine | 334/06 | 2007 |
| liposomal cytarabine | 164/05 | 2007 |
| omalizumab | 259/06 | 2007 |
| pemetrexed | 342/07 | 2007 |
| pregabalin | 389/07 | 2007 |
| rufinamide | 416/07 | 2007 |
| sevelamer | 423/07 | 2007 |
| sodium oxybate | 246/06 | 2007 |
| standardised allergen extract of grass pollen | 367/07 | 2007 |
| sunitinib | 343/07 | 2007 |
| sunitinib | 384/07 | 2007 |
| testosterone | 398/07 | 2007 |
| topotecan | 366/07 | 2007 |
| ziconotide | 405/07 | 2007 |
| atazanavir | 520/08 | 2008 |
| botulinum | 464/08 | 2008 |
| capecitabine | 507/08 | 2008 |
| clobetasol | 434/07 | 2008 |
| dabigatran | 466/08 | 2008 |
| epoetin zeta | 467/08 | 2008 |
| fondaparinux | 439/08 | 2008 |
| levetiracetam | 396/07 | 2008 |
| levetiracetam | 395/07 | 2008 |
| Methoxy polyethylene glycol-epoetin beta | 455/08 | 2008 |
| pegylated interferon alfa 2b plus ribavirin | 488/08 | 2008 |
| rivaroxaban | 519/08 | 2008 |
| sitagliptin | 505/08 | 2008 |
| telbivudine | 438/08 | 2008 |
| tenofovir | 479/08 | 2008 |
| adalimumab | 468/08 | 2008 |
| alemtuzumab | 494/08 | 2008 |
| ambrisentan | 511/08 | 2008 |
| anidulafungin | 465/08 | 2008 |
| bivalirudin | 516/08 | 2008 |
| daptomycin | 449/08 | 2008 |
| diclofenac | 446/08 | 2008 |
| docetaxel | 481/08 | 2008 |
| fesoterodine | 480/08 | 2008 |
| fosaprepitant | 506/08 | 2008 |
| imiquimod | 385/07 | 2008 |
| levetiracetam | 397/07 | 2008 |
| lidocaine | 334/06 | 2008 |
| methylnaltrexone bromide | 518/08 | 2008 |
| micafungin | 497/08 | 2008 |
| nelarabine | 454/08 | 2008 |
| nilotinib | 440/08 | 2008 |

| | | |
|---|---|---|
| pemetrexed | 342/07 | 2008 |
| raltegravir | 461/08 | 2008 |
| rituximab | 493/08 | 2008 |
| rufinamide | 416/07 | 2008 |
| vildagliptin | 435/07 | 2008 |
| zoledronic acid | 447/08 | 2008 |
| anidulafungin | 465/08 | 2008 |
| aripiprazole | 498/08 | 2008 |
| bevacizumab | 469/08 | 2008 |
| buprenorphine | 234/06 | 2008 |
| dexrazoxane | 361/07 | 2008 |
| ferric carboxymaltose | 463/08 | 2008 |
| glucosamine | 471/08 | 2008 |
| glyceryl trinitrate | 200/05 | 2008 |
| icatibant | 476/08 | 2008 |
| lenalidomide | 441/08 | 2008 |
| maraviroc | 458/08 | 2008 |
| maraviroc | 458/08 | 2008 |
| miconazole | 517/08 | 2008 |
| paliperidone | 453/08 | 2008 |
| paricalcitol | 288/06 | 2008 |
| paricalcitol | 478/08 | 2008 |
| pegylated liposomal doxorubicin | 503/08 | 2008 |
| pemetrexed | 342/07 | 2008 |
| Rabbit anti-human thymocyte immunoglobulin | 489/08 | 2008 |
| salmeterol | 450/08 | 2008 |
| sorafenib | 482/08 | 2008 |
| standardised allergen extract of grass pollen | 367/07 | 2008 |
| teriparatide | 490/08 | 2008 |
| trabectedin | 452/08 | 2008 |
| alitretinion | 538/09 | 2009 |
| fluticasone furoate | 544/09 | 2009 |
| testosterone undecanoate | 308/06 | 2009 |
| thalidomide | 525/08 | 2009 |
| doripenem | 529/09 | 2009 |
| doripenem | 539/09 | 2009 |
| fentanyl | 510/08 | 2009 |
| lacosamide | 532/09 | 2009 |
| pregabalin | 157/05 | 2009 |
| rituximab | 540/09 | 2009 |
| sugammadex | 527/09 | 2009 |
| topotecan | 545/09 | 2009 |
| aliskiren | 462/08 | 2009 |
| aripiprazole | 498/09 | 2009 |
| betaine anhydrous | 407/07 | 2009 |
| buprenorphine transdermal patch | 234/06 | 2009 |
| cetuximab | 543/09 | 2009 |
| etonogestrel/ethinyl | 502/08 | 2009 |
| etravirine | 530/09 | 2009 |
| extended-release epidural morphine sulfate | 528/09 | 2009 |
| lapatinib | 526/09 | 2009 |
| micronised progesterone | 542/09 | 2009 |
| oxycodone/naloxone | 541/09 | 2009 |
| pemetrexed | 531/09 | 2009 |
| pemetrexed | 531/09 | 2009 |
| salmeterol | 450/08 | 2009 |
| stiripentol | 524/08 | 2009 |

**SMC Dataset: Missing Data**

Between January 2004-June 2009, the Scottish Medicines Consortium (SMC) reviewed 346 full submissions/resubmissions and made 346 funding decisions – either recommending, restricting or not recommending use of NHS resources to fund new health technologies (in this research analysis, the health technologies are restricted to pharmaceutical products).

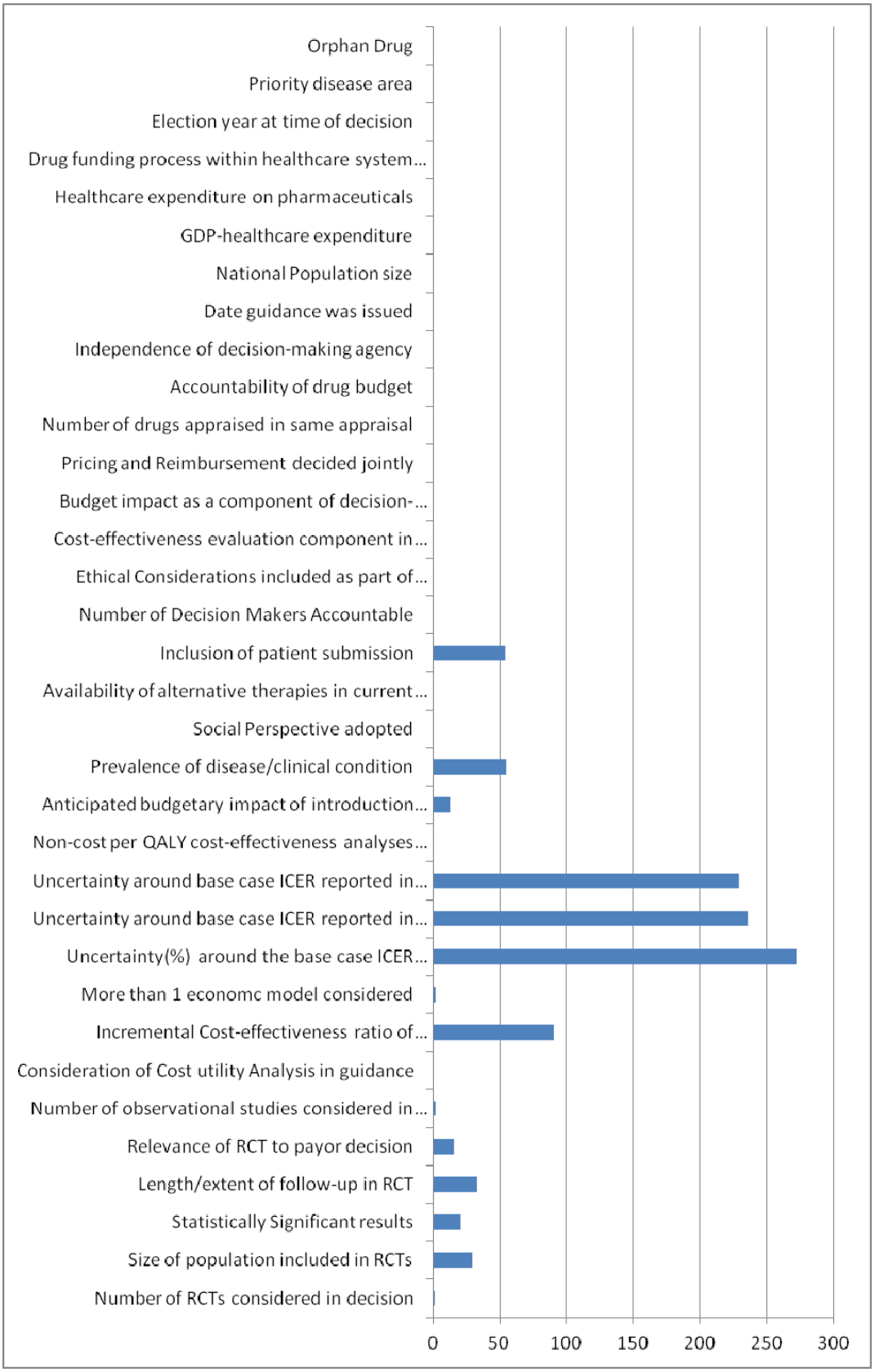
A data set of information pertaining to SMC appraisals was created collecting information on variables relating to (i) the clinical and economic characteristics of the technology under appraisal, as well as information on (ii) the process used to come to a decision, and (iii) the socio-economic context in which these decisions were made.

In order to prepare the data set for analysis, understanding the data set is important. Here, the aim is to characterise the presence of missing data across variables and decisions to inform the need for imputing missing data.

*Distribution of Missing Data within SMC data set*

In the total SMC sample, there are 10% of incomplete entries. The rate of missing entries appears similar between groups. The distribution of missing data **across each variable** was also examined. The total number of observations per variable is 288. The variables with the highest number of 'not reported' information are those related to the economic characteristics of the technology. 17 variables have no missing data. The extent of missing data was also examined **across appraisals**. The average number of 'not reported' entries per appraisal was 3 (range 0-10).

Figure B.1. SMC Distribution of missing data by variable (n=288)



**SMC Dataset – Testing of Normality Assumption**

*Method*

To determine the relevant statistical tests to use in assessing the significance of differences observed between means, it was necessary to assess whether the normality assumption was valid for the variables under consideration. This would then determine the use of parametric or non-parametric tests. For all variables, the sample is >30. It has been suggested that when analysing sample sizes of >30, even when the normality assumption is violated, parametric tests may still be performed (Pallant 2007, SPSS Survival Manual, 3rd edn, Maidenhead, OUP/McGraw-Hill). Prior to making a decision on which variables to apply parametric or non-parametric tests, the distribution for each variable will be further examined.

To test the normality assumption the following was performed for each non-categorical variable:

- Skeweness was calculated
- Kurtosis was calculated
- The standardized normal probability plot (P-P plot) was graphed

The graph and calculations are presented for each non-categorical variable and an assessment is made of whether the normality assumption is met or not. A variable will be considered to meet the normal distribution assumption if

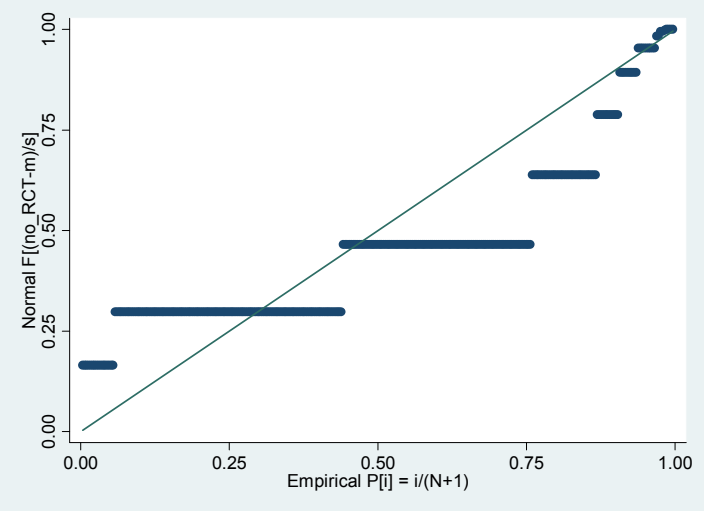
- skewness is within the range ±1
- Kurtosis value is within range ±3
- P-P plot shows distribution of dataset is approximately linear

In those circumstances where there is inconsistency between the three tests, if 2 of the tests suggest normal distribution, then it will be considered as such (but then tested using both parametric and non-parametric tests in sensitivity analyses).

*Number of RCTs considered in appraisal (No_RCT)*

Skewness 4.288793

Kurtosis 30.15434

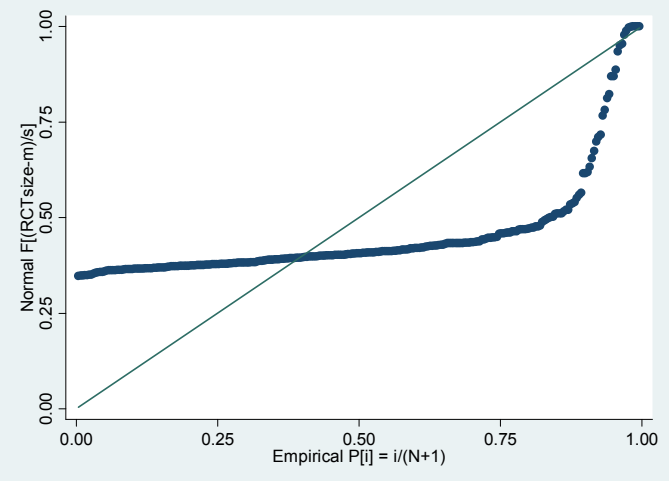



This variable is **not** normally distributed.

*Mean sample size of RCTs considered in appraisal (RCTsize)*

Skewness 6.446803

Kurtosis 51.49548




This variable is **not** normally distributed.

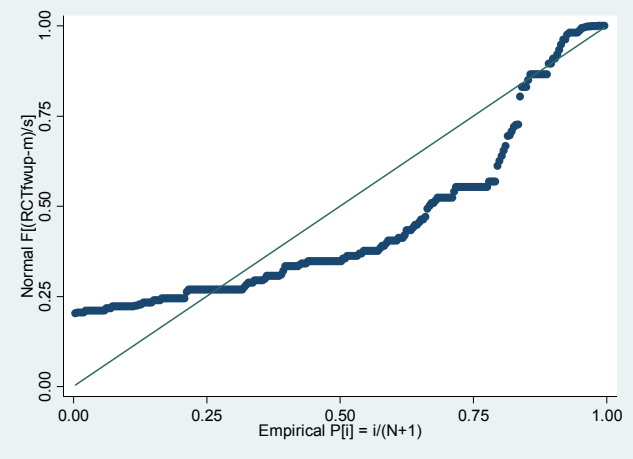
Superiority demonstrated

Normality test not performed as this is a categorical variable.

*Duration of RCT (RCTfwup)*

Skewness 2.127236

Kurtosis 7.575711

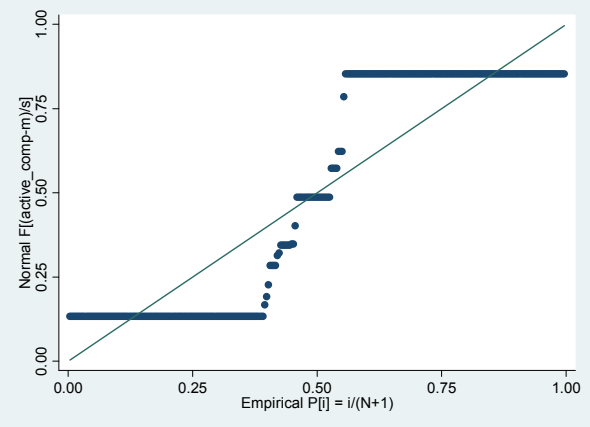



This variable is **not** normally distributed.

*Active-Comp*

Skewness -.0439793

Kurtosis 1.155245

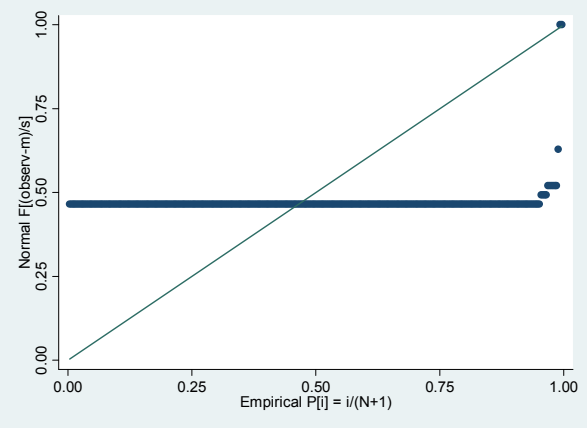



This variable **could** be considered to approximate to a normal distribution.

*Observational Studies*

Skewness 12.37259

Kurtosis 157.9963



This variable is **not** normally distributed.

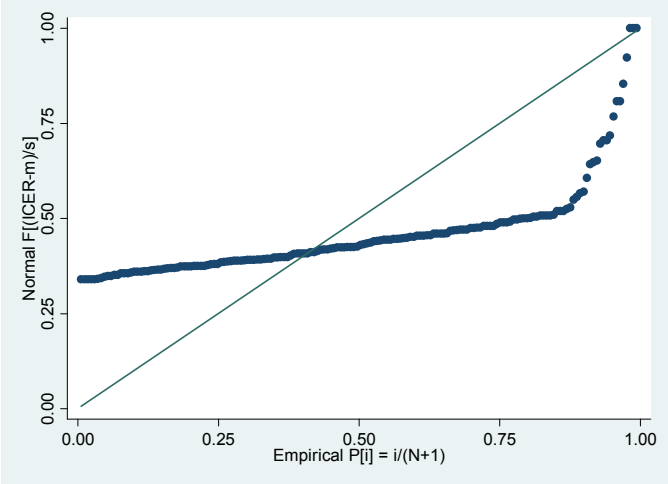
*CUA performed*

Normality test not performed as this is a categorical variable.

*ICER*

Skewness 7.445825

Kurtosis 64.31506

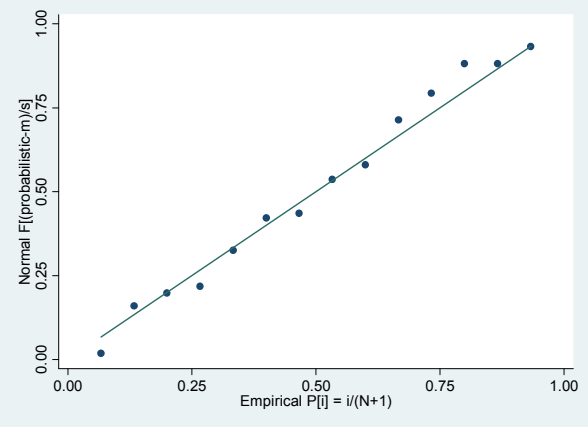



This variable is **not** normally distributed.

*ICER SA – Probabilistic*

Skewness -1.123912

Kurtosis 2.282117

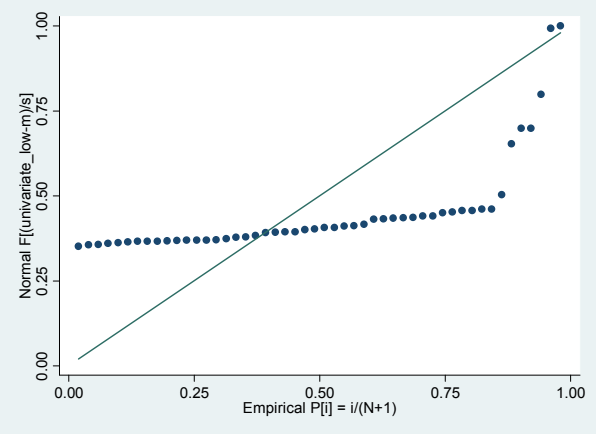



This variable **is** normally distributed.

*ICER SA – Univariate Low*

Skewness 5.214576

Kurtosis 31.50543

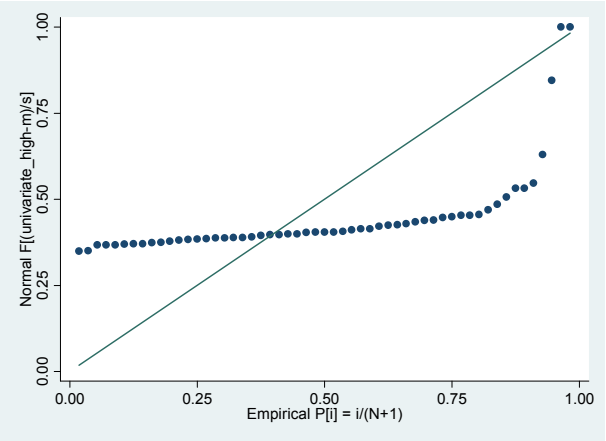



This variable is **not** normally distributed.

*ICER SA – Univariate high*

Skewness 4.881428

Kurtosis 26.57847




This variable is **not** normally distributed.

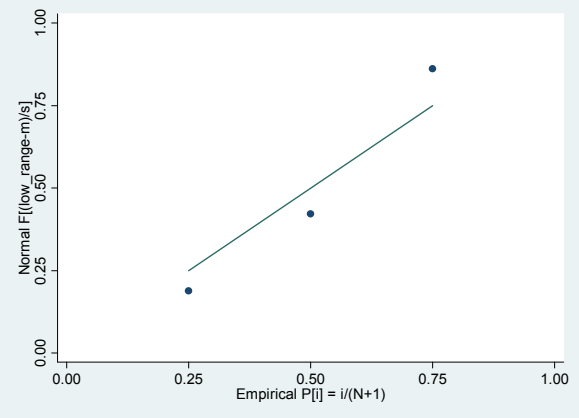
*Multiple Models submitted*

Normality test not performed as this is a categorical variable.

*Multiple Models: Range of ICERs - LOW*

Skewness .3476502

Kurtosis 1.5

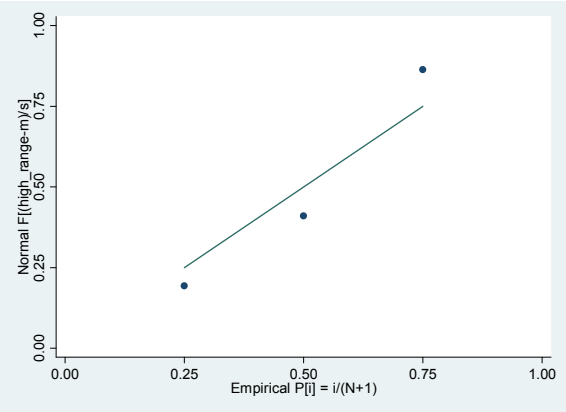



This variable **is** normally distributed.

*Multiple Models: Range of ICERs – HIGH*

Skewness .3954293

Kurtosis 1.5




This variable **is** normally distributed.

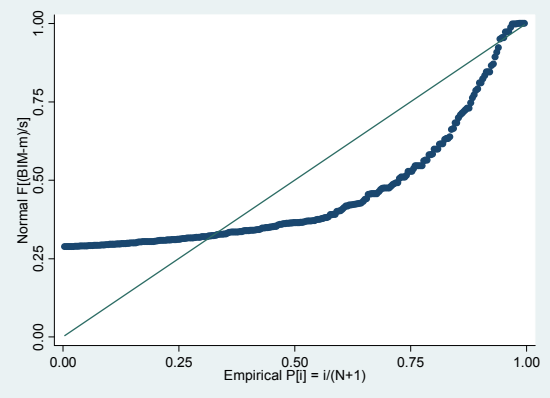
*Non_CUA submitted*

Normality test not performed as this is a categorical variable.

*BIM*

Skewness 4.359358

Kurtosis 27.93584

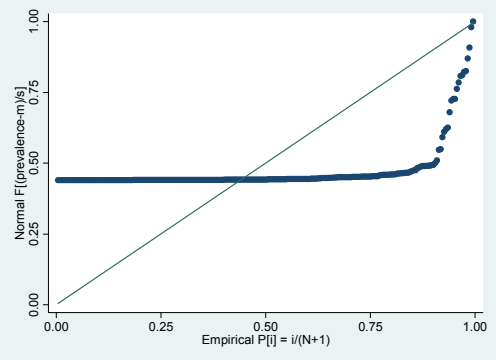



This variable is **not** normally distributed.

*Prevalence*

Skewness 13.63806

Kurtosis 198.9334




This variable is **not** normally distributed.

*Societal*

Normality test not performed as this is a categorical variable.

*Alternative Available*

Normality test not performed as this is a categorical variable.

*Group_Size*

Skewness -.4932935

Kurtosis 3.410994

This variable **could** be considered to approximate to a normal distribution.

*Cost-Effectiveness part of Process*

Normality test not performed as this is a categorical variable.

*BIM part of Process*

Normality test not performed as this is a categorical variable.

*No of Drugs appraised*

Normality test not performed as all SMC reviews consider one drug at a time - hence no variation present.

*Accountability for drug budget*

Normality test not performed as this is a categorical variable.

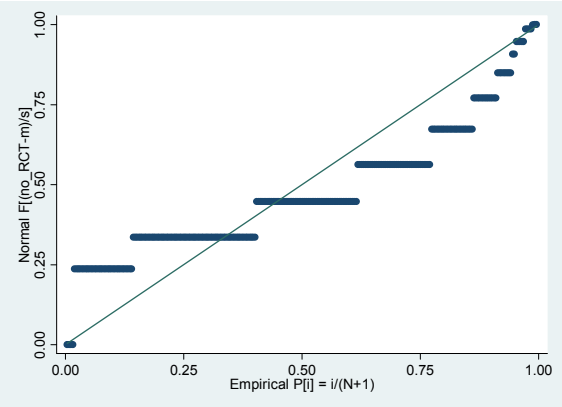
*Independence of HTA agency from MoH*

Normality test not performed as this is a categorical variable.

*Year of appraisal*

Skewness .0317761

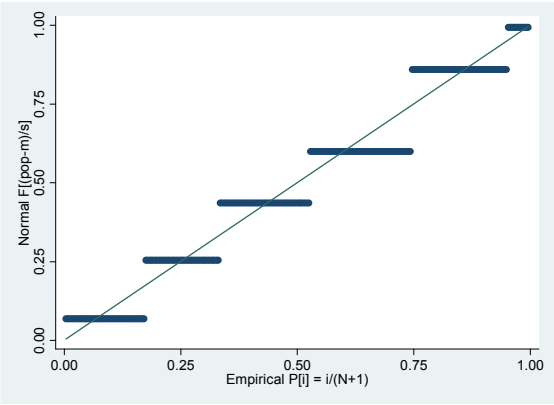
Kurtosis 1.92267




*Population under HTA remit*

Skewness .1465455

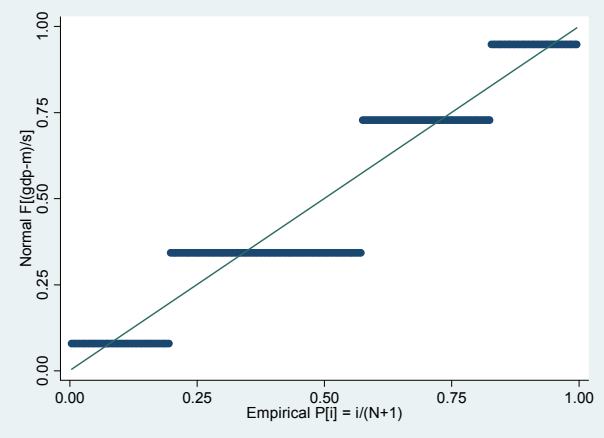
Kurtosis 1.770313




*% of GDP expenditure on health*

Skewness -.6356308

Kurtosis 1.785118




*Drug expenditure per person*

Skewness .0882442

Kurtosis 1.967974

*Funding mechanism – centralised or decentralised*

Normality test not performed as this is a categorical variable.

*Election year*

Normality test not performed as this is a categorical variable.

*Priority Disease Area*

Normality test not performed as this is a categorical variable.

*BNF Category*

Normality test not performed as this is a categorical variable.

**SMC dataset: descriptive statistics**

| Variable | Chi2 | ANOVA | T-Test (Rec vs Res) | T-Test (Res vs NR) | Kruskal-Wallis |
|---|---|---|---|---|---|
| Number of RCTs considered in decision | | 0.5898 | 0.8501 | 0.6054 | 0.3269 |
| Size of population included in RCTs | | 0.5156 | 0.7663 | 0.0048 | 0.0016 |
| Statistically Significant results - yes | 0.452 | | | | |
| no | 0.851 | | | | |
| inconsistent | 0.068 | | | | |
| Length/extent of follow-up in RCT (weeks) | | 0.4304 | 0.4082 | 0.026 | 0.023 |
| Use of Active Comparator in RCT | | 0.0032 | 0.0788 | 0.0887 | 0.0153 |
| Number of observational studies considered in guidance | | 0.0962 | 0.3534 | 0.2185 | 0.5781 |
| Consideration of Cost Utility Analysis in guidance | 0.345 | | | | |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | | 0.1312 | 0.1419 | 0.1755 | 0.0001 |
| More than one CUA submitted | 0.652 | | | | |
| If More than one CUA submitted - low | | - | - | - | 0.2207 |

| | | | | | |
|---|---|---|---|---|---|
| range | | | | | |
| If More than one CUA submitted - high range | | - | - | - | 0.2207 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | | 0.6072 | 0.9909 | 0.0176 | 0.0713 |
| Uncertainty around base case ICER reported in submission (univariate) Low | | - | 0.4548 | 0.0509 | 0.0015 |
| Uncertainty around base case ICER reported in submission (univariate) High | | 0.1901 | 0.4605 | 0.0365 | 0.0001 |
| Non-CUA analyses submitted | 0.183 | | | | |
| Potential budgetary impact (million) | | 0.8767 | 0.1941 | 0.1695 | 0.6868 |
| Prevalence of disease/clinical condition | | 0.0152 | 0.1114 | 0.3111 | 0.017 |
| Societal Perspective adopted | 0.401 | | | | |
| Availability of alternative therapies in current treatment setting. | 0.031 | | | | |
| Inclusion of patient submission | 0.105 | | | | |
| Number of Decision Makers Accountable | | 0.5704 | 0.8325 | 0.6133 | 0.8973 |
| Cost-effectiveness evaluation component in process | - | | | | |
| Budget impact as a component of decision-making process | - | | | | |
| Price of technology known during appraisal | - | | | | |
| Number of drugs appraised in same appraisal | | - | - | - | 1 |
| Accountability of drug budget | - | | | | |
| Independence of decision-making agency | - | | | | |
| Date guidance was issued | | 0.2778 | 0.5807 | 0.4118 | 0.7234 |
| Population size – Agency coverage (millions) | | 0.2778 | 0.5605 | 0.4363 | 0.4254 |
| GDP-healthcare expenditure | | 0.2778 | 0.9852 | 0.1894 | 0.7234 |
| Healthcare expenditure on pharmaceuticals | | 0.2778 | 0.6243 | 0.3177 | 0.4254 |
| Election year at time of decision | 0.906 | | | | |
| Priority disease area | 0.792 | | | | |
| Orphan Designated | 0.023 | | | | |
| Proportion of Advice following Full submission | 0.492 | | | | |
| BNF1 cardiovascular system | 0.201 | | | | |
| BNF2 central nervous system | 0.186 | | | | |
| BNF3 ear, nose and oropharynx | 0.114 | | | | |
| BNF4 endocrine system | 0.689 | | | | |
| BNF5 eye | 0.078 | | | | |
| BNF6 gastro-intestinal system | 0.028 | | | | |
| BNF7 infections | 0.007 | | | | |
| BNF8 malignant disease and immunosuppression | 0.679 | | | | |
| BNF9 musculoskeletal and joint diseases | 0.272 | | | | |
| BNF10 nutrition and blood | 0.152 | | | | |
| BNF11 obstetrics, gynaecology, and urinary-tract disorders | 0.395 | | | | |
| BNF12 respiratory system | 0.455 | | | | |
| BNF13 skin | 0.127 | | | | |

**SMC dataset: preliminary multivariate model**

The pseudo R-squared for this model was 0.1446, which suggests that it explains 15% of the variability observed in SMC coverage decisions.

**Multivariate analysis of SMC coverage decisions 2004-2009: preliminary model**

| | Coefficient | P value | 95% CI | |
|---|---|---|---|---|
| Restricted Technologies | | | | |
| RCT size | 0.1762491 | 0.152 | -0.06505 | 0.4175436 |
| RCT duration of follow-up | -0.0049669 | 0.14 | -0.01156 | 0.0016277 |
| Use of active comparator in RCT | -0.216627 | 0.661 | -1.18434 | 0.7510889 |
| Lack of clinical superiority in RCT | -3.39E-01 | 0.506 | -1.34E+00 | 0.6598822 |
| ICER | 9.88E-06 | 0.257 | -7.22E-06 | 2.70E-05 |
| Probability of ICER below £30,000 threshold | 0.0001091 | 0.195 | -5.6E-05 | 0.0002739 |
| Disease prevalence | -0.0000364 | 0.023 | -6.8E-05 | -4.96E-06 |
| Presence of alternative therapy | -0.4746392 | 0.479 | -1.78932 | 0.8400456 |
| Orphan designation status | 0.0679678 | 0.935 | -1.55539 | 1.691327 |
| Infectious Diseases | 0.1210854 | 0.827 | -0.96544 | 1.207608 |
| Musculoskeletal and joint diseases | 21.42581 | 0.000 | 20.01821 | 22.83341 |
| Obstetrics/gynaecology& urinary-tract disorders | 21.90404 | 0.000 | 19.93859 | 23.8695 |
| Skin diseases | -0.4743365 | 0.523 | -1.92943 | 0.9807586 |
| RCT size not available | 0.9686335 | 0.297 | -0.85345 | 2.790714 |
| RCT follow-up not available | -0.1260203 | 0.884 | -1.825 | 1.572959 |
| Use of active comparator not available | -2.151319 | 0.164 | -5.17925 | 0.8766147 |
| ICER not available | 0.2078835 | 0.630 | -0.63691 | 1.05268 |
| Prevalence not available | 0.4113592 | 0.404 | -0.5548 | 1.37752 |
| Constant | 1.195231 | 0.099 | -0.22414 | 2.614598 |
| Not Recommended Technologies | | | | |
| RCT size | -0.4019133 | 0.055 | -0.8122 | 0.0083758 |
| RCT duration of follow-up | -0.0100672 | 0.014 | -0.01807 | -0.0020671 |
| Use of active comparator in RCT | -0.5402572 | 0.275 | -1.51E+00 | 0.4306352 |
| Lack of clinical superiority in RCT | -5.41E-01 | 0.295 | -1.55506 | 4.72E-01 |
| ICER | 8.49E-06 | 0.342 | -9.01E-06 | 0.000026 |
| Probability of ICER below 30,000 threshold | -0.0003433 | 0.054 | -0.00069 | 6.26E-06 |
| Disease prevalence | -1.46E-06 | 0.852 | -1.68E-05 | 0.0000139 |
| Presence of alternative therapy | -0.6815033 | 0.307 | -1.98947 | 6.26E-01 |
| Orphan designation status | 0.1081514 | 0.892 | -1.45287 | 1.669176 |
| Infectious Diseases | -1.890128 | 0.006 | -3.23292 | -0.5473376 |
| Musculoskeletal and joint diseases | 20.37818 | . | . | . |
| Obstetrics/gynaecology& urinary-tract disorders | 20.6625 | . | . | . |
| Skin diseases | -2.123698 | 0.022 | -3.94227 | -0.3051237 |
| RCT size not available | 1.142295 | 0.256 | -0.82799 | 3.112579 |
| RCT follow-up not available | 0.7903202 | 0.327 | -0.78854 | 2.369185 |
| Use of active comparator not available | -2.762684 | 0.062 | -5.66286 | 0.1374896 |
| ICER not available | 0.2390104 | 0.58 | -0.60651 | 1.084535 |
| Prevalence not available | -0.0428182 | 0.932 | -1.01994 | 0.9343067 |
| Constant | 2.92186 | 0.000 | 1.470452 | 4.373268 |

Note: Recommended technologies are the reference case

**Sensitivity analyses**

The use of a categorical rather than continuous ICER variable was tested. This was done to be able to include in the analysis dominated technologies which were excluded from the base-case analysis as they were challenging to quantify on a continuous scale. The categorical ICER variable was created by generating a binary variable which recorded whether the ICER was above or below £30,000. This ICER threshold value was selected on the assumption that it represents the value upon which a technology is deemed to be cost-effective versus the comparator by the SMC. To test how the new ICER variable impacted on the regression output, the continuous ICER variable in the above regression model was replaced with the binary ICER variable. The results are shown in Table 5.7. When compared to the base case regression output, the results of this sensitivity analysis suggest the model using a categorical ICER variable rather than a continuous ICER variable yields similar results. The pseudo R-squared is similar (0.1239 vs. 0.1103 in the base case), and the categorical ICER variable maintains the statistically significant effect on the log odds of a restriction and non-recommendation relative to a recommendation.

**Sensitivity Analysis 1. Multivariate analysis of SMC coverage decisions 2004-2009: sensivitiy analysis using categorical ICER variable**

| | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Restricted Technologies | | | | |
| RCT size | 1.34E-05 | 0.841 | -0.00012 | 0.000144 |
| RCT duration of follow-up | -0.00301 | 0.315 | -0.00889 | 0.002864 |
| ICER | -1.36422 | 0.092 | -2.95262 | 0.224181 |
| Infectious Diseases | -0.58987 | 0.259 | -1.61407 | 0.434338 |
| Skin Diseases | -0.74685 | 0.324 | -2.2303 | 0.736596 |
| Disease prevalence | -1.6E-05 | 0.097 | -3.4E-05 | 2.84E-06 |
| Presence of alternative therapy | -0.84396 | 0.225 | -2.20669 | 0.518771 |
| Constant | 2.9919 | 0.003 | 0.984899 | 4.998901 |
| Not recommended technologies | | | | |
| RCT size | -0.00033 | 0.073 | -0.0007 | 3.09E-05 |
| RCT duration of follow-up | -0.00882 | 0.022 | -0.01638 | -0.00126 |
| ICER | -2.35043 | 0.003 | -3.91486 | -0.786 |
| Infectious Diseases | -2.16812 | 0.001 | -3.49508 | -0.84116 |
| Skin Diseases | -1.85797 | 0.045 | -3.67625 | -0.03968 |
| Disease prevalence | -2E-05 | 0.045 | -3.9E-05 | -4.78E-07 |
| Presence of alternative therapy | -1.45758 | 0.039 | -2.83982 | -0.07534 |
| Constant | 5.109186 | 0 | 3.089329 | 7.129043 |

Note: Recommended technologies are the reference case

# C. Chapter 6 Appendices

## CFH coverage decisions 2004-2009: List of Technology Appraisals included for analysis

| Technology Appraised | Date guidance was issued |
|---|---|
| abacavir/lamivudine | 2005 |
| abatacept | 2007 |
| adalimumab | 2005 |
| adalimumab | 2006 |
| adalimumab | 2007 |
| adalimumab | 2007 |
| adalimumab | 2008 |
| adalimumab | 2008 |
| adalimumab | 2008 |
| adapaleen | 2005 |
| agalsidase alfa | 2007 |
| agalsidase beta | 2007 |
| alemtuzumab | 2006 |
| alemtuzumab | 2008 |
| alendroninezuur/colecalciferol | 2005 |
| alfa1-proteinaseremmer | 2007 |
| aliskiren | 2007 |
| allergenen | 2008 |
| allergenen | 2008 |
| allergenen | 2008 |
| allergenen | 2008 |
| allergenen | 2008 |
| ambrisentan | 2008 |
| anagrelide | 2005 |
| anakinra | 2006 |
| anakinra | 2006 |
| anidulafungine | 2008 |
| aprepitant | 2004 |
| aripiprazol | 2004 |
| atazanavir | 2004 |
| betaïneanhydraat | 2007 |
| bevacizumab | 2005 |
| bevacizumab | 2007 |
| bevacizumab | 2007 |
| bevacizumab | 2008 |
| bortezomib | 2005 |
| bortezomib | 2007 |
| bortezomib | 2008 |
| bosentan | 2007 |
| buprenorfine | 2007 |
| buprenorfine/naloxon | 2009 |
| bupropion | 2007 |
| carglumaatzuur | 2006 |
| caspofungine | 2008 |
| celecoxib | 2004 |
| cetuximab | 2007 |
| cetuximab | 2007 |
| ciclesonide | 2005 |
| cinacalcet | 2005 |
| ciprofloxacine | 2005 |
| clopidogrel | 2004 |
| clopidogrel | 2004 |

| | |
|---|---|
| clopidogrel | 2007 |
| clopidogrel | 2009 |
| colesevelam | 2007 |
| colistine | 2005 |
| cytarabine | 2007 |
| dabigatran | 2008 |
| darbepoëtine | 2006 |
| darbepoëtine-alfa | 2007 |
| darifenacine | 2005 |
| darunavir | 2007 |
| dasatinib | 2007 |
| deferasirox | 2006 |
| dorzolamide | 2006 |
| dorzolamide/timolol | 2006 |
| drotrecogin-alfa | 2006 |
| duloxetine | 2004 |
| duloxetine | 2005 |
| eculizumab | 2008 |
| efalizumab | 2005 |
| emtricitabine | 2004 |
| entecavir | 2006 |
| epinastine | 2004 |
| eplerenon | 2004 |
| epoëtine | 2006 |
| erytromycine-zinkcomplex | 2005 |
| escitalopram | 2004 |
| escitalopram | 2004 |
| escitalopram | 2004 |
| estradiol/drospirenon | 2004 |
| etanercept | 2004 |
| etanercept | 2004 |
| etanercept | 2008 |
| etanercept | 2008 |
| etanercept | 2008 |
| ethinylestradiol | 2004 |
| etravirine | 2008 |
| everolimus | 2004 |
| exenatide | 2007 |
| fentanyl oromucosaal | 2006 |
| fesoterodine | 2008 |
| fluticasonfuroaat | 2008 |
| fosamprenavir | 2004 |
| fulvestrant | 2004 |
| fumaarzuuresters | 2004 |
| fumaarzuuresters | 2004 |
| fumaarzuuresters | 2004 |
| galsulfase | 2007 |
| gamma hydroxyboterzuur | 2006 |
| gepegyleerd liposomaal doxorubicine | 2005 |
| gepegyleerd liposomaal doxorubicine | 2005 |
| gepegyleerd liposomaal doxorubicine | 2005 |
| gliclazide | 2005 |
| humaan papillomavirusvaccin | 2007 |
| humaan papillomavirusvaccin | 2009 |
| ibandroninezuur | 2004 |
| ibandroninezuur | 2005 |
| ibandroninezuur i.v. | 2006 |
| ibandroninezuur i.v. | 2006 |
| ibandroninezuur i.v. | 2006 |
| ibritumomab | 2006 |

| | |
|---|---|
| ijzerdextraan | 2004 |
| iloprost | 2008 |
| imiquimod | 2005 |
| infliximab | 2005 |
| infliximab | 2005 |
| infliximab | 2006 |
| infliximab | 2006 |
| infliximab | 2006 |
| infliximab | 2007 |
| infliximab | 2007 |
| insuline detemir | 2004 |
| insuline glulisine | 2005 |
| irinotecan | 2005 |
| isosorbidedinitraat | 2007 |
| ivabradine | 2006 |
| ivabradine | 2007 |
| lacosamide | 2009 |
| lanthaancarbonaat | 2006 |
| lapatinib | 2009 |
| lenalidomide | 2007 |
| levetiracetam | 2005 |
| levetiracetam | 2008 |
| levodopa/carbidopa | 2005 |
| levodopa/carbidopa/entacapone | 2004 |
| lidocaïnepleister | 2008 |
| lomustine | 2006 |
| lumiracoxib | 2007 |
| macrogol/elektrolyten | 2007 |
| maraviroc | 2007 |
| melatonine | 2004 |
| melatonine | 2005 |
| melatonine | 2008 |
| melatonine | 2008 |
| memantine | 2004 |
| memantine | 2007 |
| menopauzegonadotrofine | 2004 |
| methoxypolyethyleenglycolepoëtine beta | 2007 |
| methylaminolevulinaat | 2008 |
| methylaminolevulinaat | 2008 |
| methylaminolevulinaat | 2008 |
| methylnaltrexon | 2008 |
| mexiletine | 2006 |
| miconazol | 2009 |
| miglustat | 2009 |
| mitotane | 2006 |
| mycofenolaatmofetil | 2006 |
| mycofenolaatmofetil | 2006 |
| mycofenolaatmofetil | 2007 |
| mycofenolaatmofetil | 2008 |
| mycofenolzuur | 2004 |
| myrtol | 2005 |
| natalizumab | 2006 |
| nepafenac | 2008 |
| nicotinezuur | 2004 |
| nilotinib | 2008 |
| ofloxacine | 2006 |
| omega-3-vetzuren | 2004 |
| orlistat | 2004 |
| oseltamivir | 2005 |
| oxybutynine | 2005 |

| | |
|---|---|
| palifermin | 2006 |
| paliperidon | 2007 |
| panitumumab | 2008 |
| parathyroïd hormoon | 2006 |
| paricalcitol | 2008 |
| pegaptanib | 2006 |
| peginterferon-alfa 2 | 2008 |
| pemetrexed | 2005 |
| pemetrexed | 2005 |
| pimecrolimus | 2004 |
| pimecrolimus | 2004 |
| pioglitazon | 2007 |
| posaconazol | 2006 |
| posaconazol | 2007 |
| pregabaline | 2004 |
| pregabaline | 2004 |
| quetiapine | 2008 |
| quetiapine | 2008 |
| raltegravir | 2008 |
| ranibizumab | 2007 |
| rasagiline | 2006 |
| rimonabant | 2007 |
| rituximab | 2006 |
| rituximab | 2006 |
| rituximab | 2006 |
| rivaroxaban | 2009 |
| rivaroxaban | 2009 |
| rivastigmine | 2006 |
| rosiglitazon | 2007 |
| rosiglitazon/metformine | 2004 |
| rotigotine | 2006 |
| rotigotine | 2007 |
| sapropterin | 2009 |
| sertindol | 2006 |
| sildenafil | 2006 |
| sitagliptine | 2007 |
| sitagliptine | 2008 |
| sitaxentan | 2007 |
| solifenacine | 2004 |
| sorafenib | 2006 |
| sorafenib | 2008 |
| sorafenib | 2009 |
| strontiumranelaat | 2005 |
| sunitinib | 2006 |
| sunitinib | 2006 |
| sunitinib | 2007 |
| tacrolimus | 2006 |
| telbivudine | 2007 |
| temoporfine | 2008 |
| temoporfine | 2008 |
| temsirolimus | 2008 |
| teriparatide | 2004 |
| testosteron | 2007 |
| testosteron gel | 2004 |
| testosteronpleister | 2008 |
| tetrabenazine | 2007 |
| tetrabenazine | 2009 |
| thalidomide | 2007 |
| thalidomide | 2007 |
| thalidomide | 2007 |

| | |
|---|---|
| timolol/brimonidine | 2006 |
| tipranavir | 2006 |
| tolcapon | 2005 |
| tolcapon | 2006 |
| topotecan | 2008 |
| trabectedin | 2008 |
| trastuzumab | 2005 |
| treprostinil | 2006 |
| tretinoïne | 2005 |
| urofollitropine | 2007 |
| urofollitropine | 2007 |
| varenicline | 2008 |
| vildagliptine | 2008 |
| vinorelbine | 2005 |
| vinorelbine | 2005 |
| voriconazol | 2008 |
| xycodon | 2007 |
| zileuton | 2008 |
| zinc acetate | 2005 |
| zoledroninezuur | 2004 |
| zoledroninezuur | 2004 |
| zoledroninezuur | 2005 |
| zoledroninezuur | 2008 |
| zonisamide | 2007 |

**CFH Dataset: Missing Data**

Between January 2004-June 2009, the Commissie Farmaceutische Hulp (CFH) reviewed 256 submissions and made specific funding decisions – either recommending, restricting or not recommending use of Dutch health care resources to fund new health technologies (in this research analysis, the health technologies are restricted to pharmaceutical products).

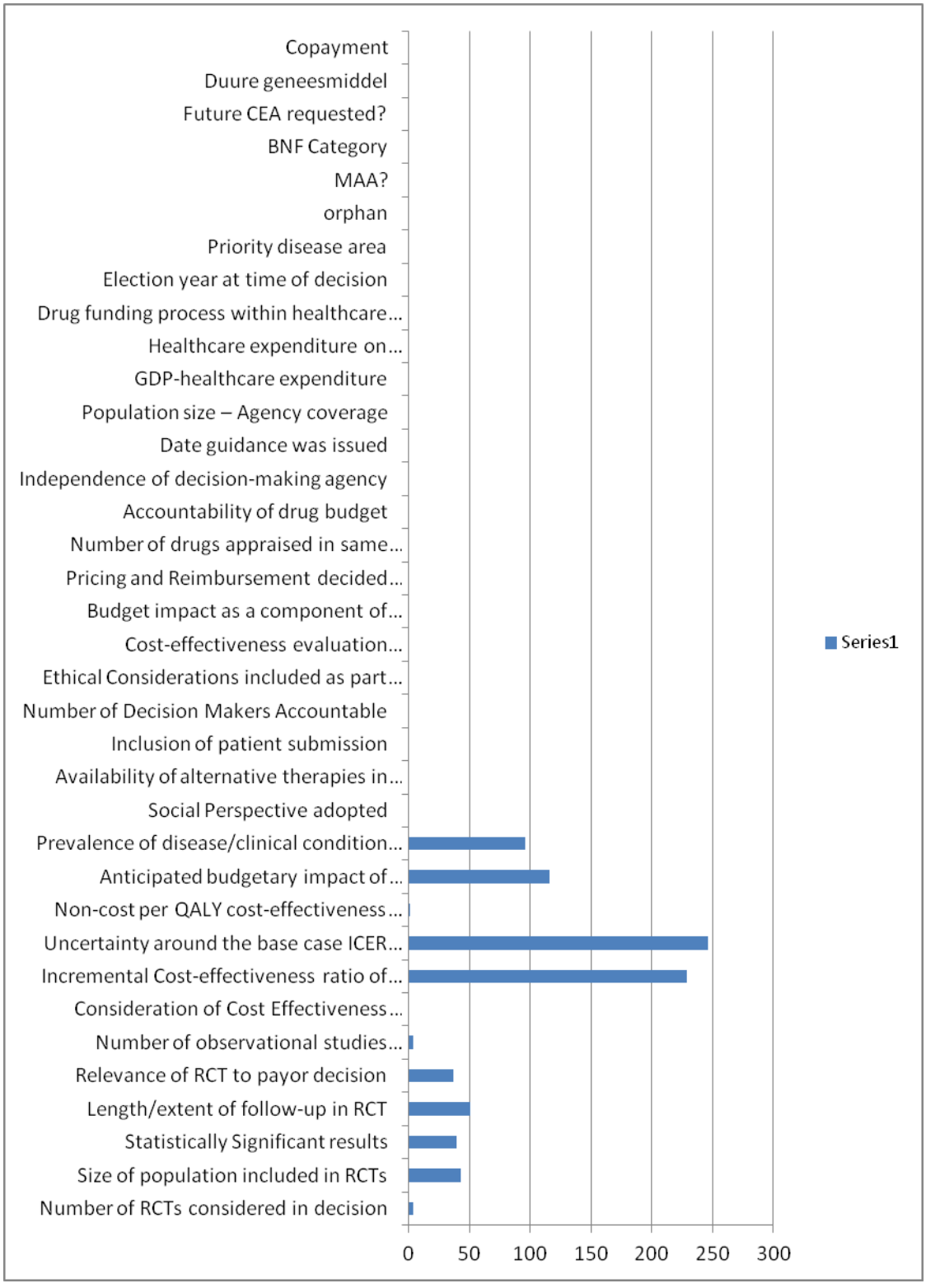
A data set of information pertaining to CFH appraisals was created collecting information on variables relating to (i) the clinical and economic characteristics of the technology under appraisal, as well as information on (ii) the process used to come to a decision, and (iii) the socio-economic context in which these decisions were made. In order to prepare the data set for analysis, understanding the data set is important. Here, the aim is to characterise the presence of missing data across variables and decisions to inform the need for imputing missing data.

*Distribution of Missing Data within CFH data set*

In the total CFH sample, there are 9% of incomplete entries. The rate of missing entries is similar across decision outcomes. The distribution of missing data **across each variable** was examined. The total number of observations per variable is 256. The variables with the highest number missing information are those related to the prevalence and budget impact of the technology, as well as the incremental cost-effectiveness ratio and related variables. This is linked partially to the fact that until 2006, cost-effectiveness models were not a formal component of the CFH review process. More than half of all variables (n=26) have no missing data.

The extent of missing data was also examined **across appraisals.** The level of missing variables per appraisal ranged from 0 – 8, mean number of missing entries per appraisal was 3.

Figure C.1. CVZ Distribution of missing data by variable (n=256)



**CFH Dataset – Testing of Normality Assumption**

Method

To determine the relevant statistical tests to use in assessing the significance of differences observed between means, it was necessary to assess whether the normality assumption was valid for the variables under consideration. This would then determine the use of parametric or non-parametric tests. For all variables, the sample is >30. It has been suggested that when analysing sample sizes of >30, even when the normality assumption is violated, parametric tests may still be performed (Pallant 2007, SPSS Survival Manual, 3rd edn, Maidenhead, OUP/McGraw-Hill). Prior to making a decision on which variables to apply parametric or non-parametric tests, the distribution for each variable will be further examined.

To test the normality assumption the following was performed for each non-categorical variable:

- Skeweness was calculated
- Kurtosis was calculated
- The standardized normal probability plot (P-P plot) was graphed

The graph and calculations are presented for each non-categorical variable and an assessment is made of whether the normality assumption is met or not. A variable will be considered to meet the normal distribution assumption if

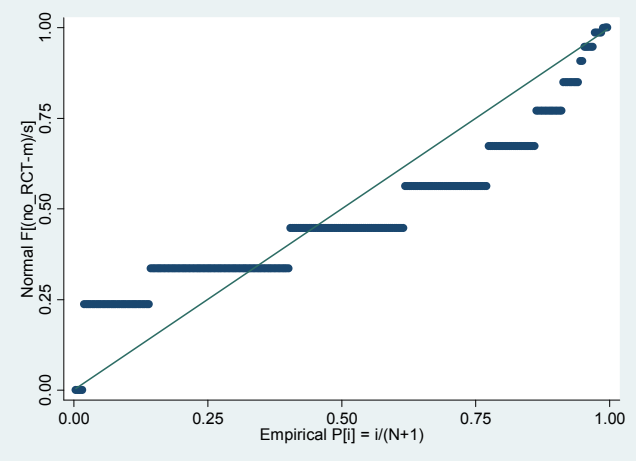
- skewness is within the range ±1
- Kurtosis value is within range ±3
- P-P plot shows distribution of dataset is approximately linear

In those circumstances where there is inconsistency between the three tests, if 2 of the tests suggest normal distribution, then it will be considered as such (but then tested using both parametric and non-parametric tests in sensitivity analyses).

**Number of RCTs considered in appraisal (No_RCT)**

Skewness 5.058098

Kurtosis 42.76271

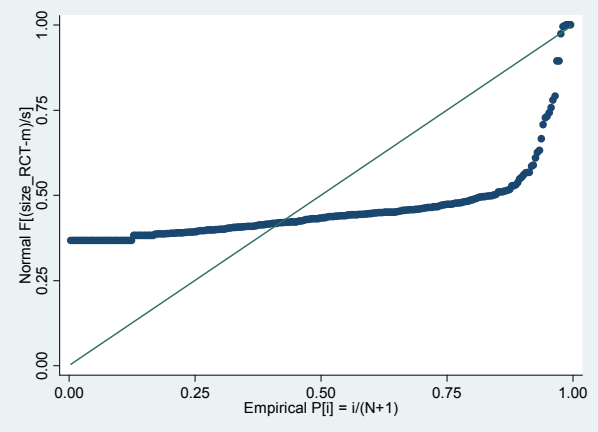



This variable is **not** normally distributed.

**Mean sample size of RCTs considered in appraisal (RCTsize)**

Skewness 7.285689

Kurtosis 60.86852




This variable is **not** normally distributed.

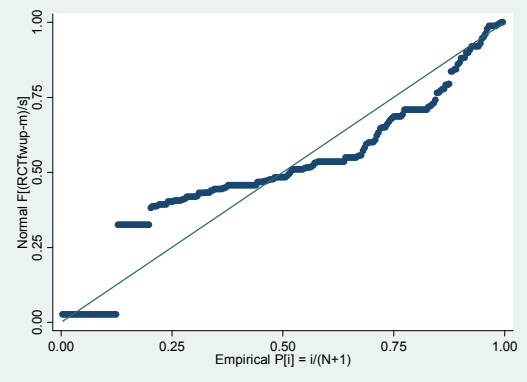
**Superiority demonstrated**

Normality test not performed as this is a categorical variable.

**Duration of RCT (RCTfwup)**

Skewness 2.550881

Kurtosis 12.20497




This variable is normally distributed.

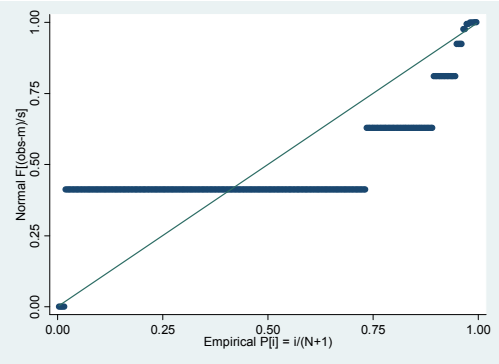
**Active-Comp**

Categorical variable.

**Observational Studies**

Skewness 4.508569

Kurtosis 29.06404




This variable is **not** normally distributed.

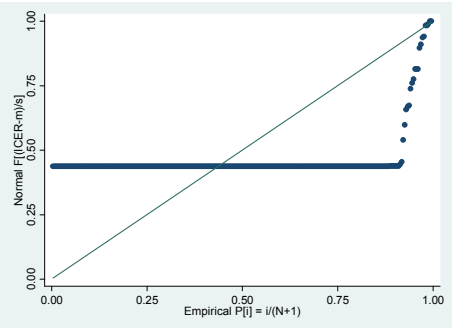
**CUA performed**

Normality test not performed as this is a categorical variable.

**ICER**

Skewness 3.950522

Kurtosis 18.84594

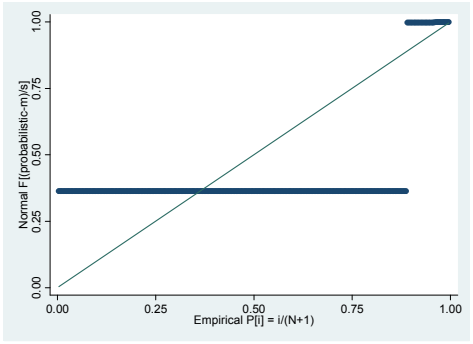


This variable is **not** normally distributed.

**ICER SA – Probabilistic**

Skewness -1.123912

Kurtosis 2.282117

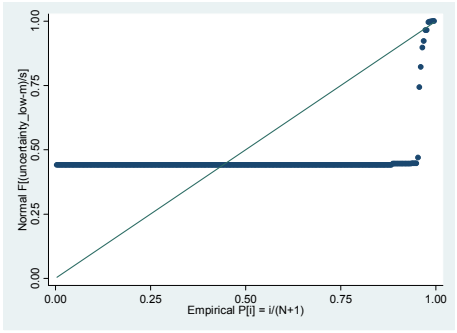



This variable **is** normally distributed.

**ICER SA – Univariate Low**

Skewness 5.214576

Kurtosis 31.50543

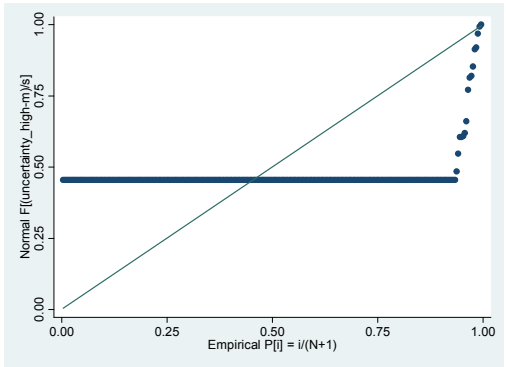



This variable is **not** normally distributed.

**ICER SA – Univariate high**

Skewness 4.881428

Kurtosis 26.57847




This variable is **not** normally distributed.

**Multiple Models submitted**

Normality test not performed as this is a categorical variable.

**Multiple Models: Range of ICERs – LOW / HIGH**

Normality test not performed as there are only 2 observations for this particular variable.

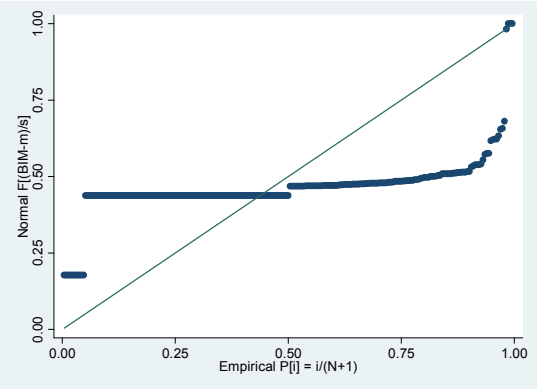
**Non_CUA submitted**

Normality test not performed as this is a categorical variable.

**BIM**

Skewness 6.388764

Kurtosis 44.28053

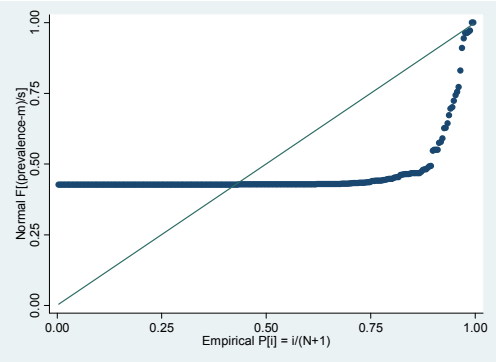



This variable is **not** normally distributed.

**Prevalence**

Skewness 7.605384

Kurtosis 64.17695




This variable is **not** normally distributed.

**Alternative Available**

Normality test not performed as this is a categorical variable.

**Cost-Effectiveness part of Process**

Normality test not performed as this is a categorical variable.

**BIM part of Process**

Normality test not performed as this is a categorical variable.

**Joint reimbursement & pricing decision**

Normality test not performed as this is a categorical variable.

**Accountability for drug budget**

Normality test not performed as this is a categorical variable.

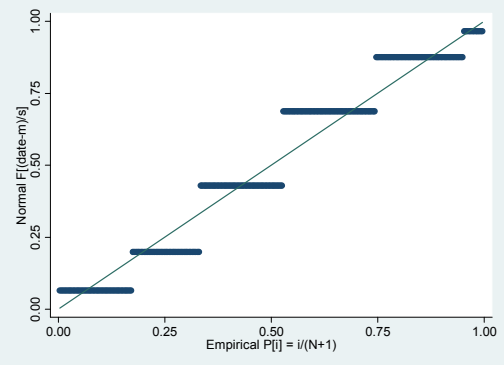
**Independence of HTA agency from MoH**

Normality test not performed as this is a categorical variable.

**Year of appraisal**

Skewness -.0679571

Kurtosis 1.887242

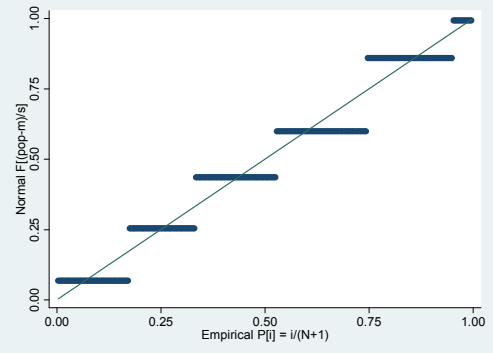



Perhaps this is actually a categorical variable?

**Population under HTA remit**

Skewness .3594944

Kurtosis 2.930518

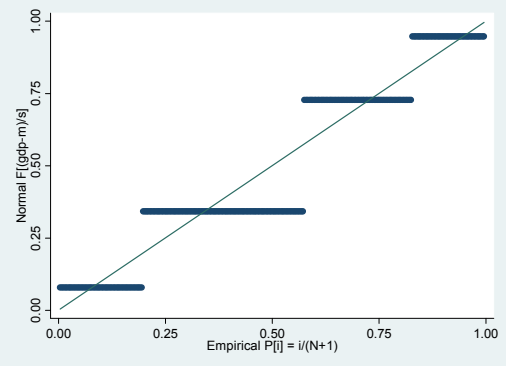



Perhaps this is actually a categorical variable?

**% of GDP expenditure on health**

Skewness .1992813

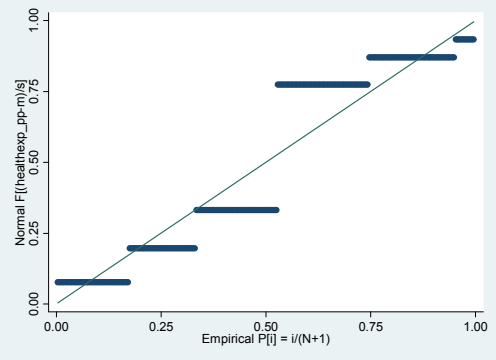
Kurtosis 2.01775




**Drug expenditure per person**

Skewness 2.01775

Kurtosis 1.457256




**Funding mechanism – centralised or decentralised**

Normality test not performed as this is a categorical variable.

**Election year**

Normality test not performed as this is a categorical variable.

**Priority Disease Area**

Normality test not performed as this is a categorical variable.

**BNF Category**

Normality test not performed as this is a categorical variable.

**CFH dataset: descriptive statistics**

| | Chi2 | ANOVA | T-Test (Rec vs Res) | T-Test (Res vs NR) | Kruskal-Wallis |
|---|---|---|---|---|---|
| Number of RCTs considered in decision | | 0.039 | 0.0595 | 0.1987 | 0.0912 |
| Size of population included in RCTs | | 0.1027 | 0.0357 | 0.2878 | 0.3042 |
| Statistically Significant results - yes | 0.046 | | | | |
| no | 0.003 | | | | |
| inconsistent | 0.922 | | | | |
| Length/extent of follow-up in RCT (weeks) | | 0.2398 | 0.0467 | 0.1109 | 0.2563 |
| Use of Active Comparator in RCT | | 0.0836 | 0.3306 | 0.0043 | 0.0081 |
| Number of observational studies considered in guidance | | 0.8655 | 0.9032 | 0.7112 | 0.7948 |
| Consideration of Cost Utility Analysis in guidance | 0.196 | | | | |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | | 0.5432 | 0.4799 | 0.9323 | 0.5335 |
| More than one CUA submitted | 0.23 | | | | |
| If More than one CUA submitted - low range | | - | - | - | |
| If More than one CUA submitted - high range | | - | - | - | |
| Uncertainty around the base case ICER reported in submission (probabilistic) | | 0.5774 | 0.4387 | 0.9362 | |
| Uncertainty around base case ICER reported in submission (univariate) Low | | 0.5228 | 0.4939 | 0.6339 | 0.5538 |
| Uncertainty around base case ICER reported in submission (univariate) High | | 0.5344 | 0.4772 | 0.883 | 0.5538 |
| Prevalence of disease/clinical condition | | 0.0193 | 0.7033 | 0.0302 | 0.0027 |
| Potential budgetary impact (million) | | 0.3603 | 0.0493 | 0.9044 | 0.0224 |
| Societal Perspective adopted | 0.623 | | | | |
| Availability of alternative therapies in current treatment setting. | 0.276 | | | | |
| Inclusion of patient submission | 0.009 | | | | |
| Number of Decision Makers Accountable | | - | - | - | 1 |
| Cost-effectiveness evaluation component in process | 0.287 | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Budget impact as a component of decision-making process | - | - | - | - | |
| Price of technology known during appraisal | - | | | | |
| Number of drugs appraised in same appraisal | | - | - | - | 1 |
| Accountability of drug budget | - | - | - | - | |
| Independence of decision-making agency | - | - | - | - | |
| Date guidance was issued | | 0.1349 | 0.6864 | 0.0158 | 0.0333 |
| Population size – Agency coverage (millions) | | 0.1349 | 0.5654 | 0.0222 | 0.0333 |
| GDP-healthcare expenditure | | 0.1634 | 0.4856 | 0.3973 | 0.3365 |
| Healthcare expenditure on pharmaceuticals | | 0.1349 | 0.8511 | 0.0099 | 0.0333 |
| Election year at time of decision | 0.094 | | | | |
| Priority disease area | 0.028 | | | | |
| Orphan Designated | 0.1 | | | | |
| Technology has EU Marketing Authorisation | 0.207 | | | | |
| Future Cost-Effectiveness Analyses requested | 0.0001 | | | | |
| Expensive Drug | 0.0001 | | | | |
| Patient Copayment needed | 0.019 | | | | |
| BNF1 cardiovascular system | 0.014 | | | | |
| BNF2 central nervous system | 0.526 | | | | |
| BNF3 ear, nose and oropharynx | 0.23 | | | | |
| BNF4 endocrine system | 0.541 | | | | |
| BNF5 eye | 0.14 | | | | |
| BNF6 gastro-intestinal system | 0.752 | | | | |
| BNF7 infections | 0.351 | | | | |
| BNF9 musculoskeletal and joint diseases | 0.005 | | | | |
| BNF10 nutrition and blood | 0.471 | | | | |
| BNF11 obstetrics, gynaecology, and urinary-tract disorders | 0.036 | | | | |
| BNF12 respiratory system | 0.0001 | | | | |
| BNF13 skin | 0.382 | | | | |

## CFH preliminary multivariate model

The result of the multinomial regression yield a pseudo R-squared of 0.2276, suggesting that the model explains 23% of the variability in CFH coverage decisions.

**Multivariate analysis of CFH coverage decisions 2004-2009: preliminary model**

| **Restricted** | **Log Odds** | **P value** | **95% Confidence Interval** | |
|---|---|---|---|---|
| Number of RCTs | 0.0021636 | 0.975 | -0.1338959 | 0.138223 |
| RCT duration of follow-up | -0.0041777 | 0.435 | -0.0146733 | 0.006318 |
| Use of active comparator in RCT | -0.5027428 | 0.282 | -1.419362 | 0.4138761 |
| Demonstrated clinical superiority in RCT | -0.1743879 | 0.739 | -1.199567 | 0.8507908 |
| Inconsistently demonstrated superiority in RCT | 0.641352 | 0.212 | -0.3665792 | 1.649283 |
| Budgetary Impact | 0.0063992 | 0.243 | -0.0043358 | 0.0171342 |
| Patient Submission | 0.0931105 | 0.929 | -1.961057 | 2.147278 |
| Cancer therapy | -0.8279568 | 0.180 | -2.039338 | 0.3834244 |
| Therapies for cardiovascular diseases | 1.36122 | 0.036 | 0.0917759 | 2.630663 |
| Therapies for musculoskeletal and joint diseases | 0.9945765 | 0.150 | -0.3605296 | 2.349683 |
| Therapies for obstetrics/gynaecology/urinary-tract disorders | 2.24357 | 0.053 | -0.0255119 | 4.512652 |
| Therapies for respiratory system | -33.71069 | 1.000 | -4.99E+07 | 4.99E+07 |
| National Population | -6.96E-06 | 0.453 | -0.0000251 | 0.0000112 |
| Pharmaceutical Expenditure per patient per year | 0.0181555 | 0.495 | -0.0339831 | 0.0702941 |
| Priority Disease area | -0.4878657 | 0.316 | -1.442324 | 0.4665928 |
| Election | 0.0913282 | 0.833 | -0.7568523 | 0.9395087 |
| Lack of data on duration of RCT | 0.7001897 | 0.406 | -0.9507739 | 2.351153 |
| Lack of data on use of active comparator | -0.7064949 | 0.448 | -2.532211 | 1.119221 |
| Lack of data on budgetary impact | 0.2760691 | 0.480 | -0.4906909 | 1.042829 |
| Constant | 108.0983 | 0.453 | -174.4742 | 390.6708 |
| **Not Recommended** | **Log Odds** | **P value** | **95% Confidence Interval** | |
| Number of RCTs | -0.0529028 | 0.514 | -0.2117183 | 0.1059126 |
| RCT duration of follow-up | 0.0011882 | 0.856 | -0.0116559 | 0.0140324 |
| Use of active comparator in RCT | -2.111281 | 0.004 | -3.546022 | -0.6765408 |
| Demonstrated clinical superiority in RCT | -0.5771367 | 0.434 | -2.02204 | 0.8677663 |
| Inconsistently demonstrated superiority in RCT | 0.924818 | 0.198 | -0.4825063 | 2.332142 |
| Budgetary Impact | 0.00698 | 0.206 | -0.0038408 | 0.0178008 |
| Patient Submission | 1.812627 | 0.049 | 0.0092095 | 3.616045 |
| Cancer therapy | -1.062929 | 0.146 | -2.49712 | 0.3712627 |
| Therapies for cardiovascular diseases | 1.271848 | 0.141 | -0.4202808 | 2.963977 |
| Therapies for musculoskeletal and joint diseases | -1.540254 | 0.216 | -3.981781 | 0.9012729 |
| Therapies for obstetrics, gynaecology, and urinary-tract disorders | 2.334487 | 0.137 | -0.7455755 | 5.41455 |
| Therapies for respiratory system | 2.088872 | 0.122 | -0.5611742 | 4.738918 |
| National Population | -5.46E-06 | 0.620 | -0.0000271 | 0.0000161 |

| | | | | |
|---|---|---|---|---|
| Pharmaceutical Expenditure per patient per year | 0.0257991 | 0.428 | -0.0379471 | 0.0895454 |
| Priority Disease area | 0.3928719 | 0.539 | -0.8610197 | 1.646763 |
| Election | -0.5338336 | 0.468 | -1.976561 | 0.9088942 |
| Lack of data on duration of RCT | 2.065647 | 0.045 | 0.0442537 | 4.087041 |
| Lack of data on use of active comparator | -1.050057 | 0.377 | -3.379514 | 1.2794 |
| Lack of data on budgetary impact | 0.0992704 | 0.852 | -0.9404436 | 1.138984 |
| Constant | 80.53906 | 0.638 | -254.852 | 415.9302 |

Note: Recommended technologies are the reference case

# D. Chapter 7 Appendices

## HAS coverage decisions 2004-2009: List of Technology Appraisals included for analysis

| Drug Name | Ref HAS | Date guidance was issued |
|---|---|---|
| adalimumab | 31648 | 2004 |
| adalimumab | 31580 | 2004 |
| anastrozole | 31571 | 2004 |
| Aprepitant | 31576 | 2004 |
| aripiprazole | 31711 | 2004 |
| atazanavir | 31704 | 2004 |
| atorvastatin | 31682 | 2004 |
| bortezomib | 31654 | 2004 |
| Budesonide/eformoterol | 31498 | 2004 |
| busulfan | 31489 | 2004 |
| Calcipotriol and betamethasone | 31508 | 2004 |
| docetaxel | 31612 | 2004 |
| emtricitabine | 31564 | 2004 |
| escitalopram | 31652 | 2004 |
| esomeprazole | 31649 | 2004 |
| etanercept | 31531 | 2004 |
| fludarabine | 31547 | 2004 |
| fondaparinux | 31572 | 2004 |
| fondaparinux | 31572 | 2004 |
| fosamprenavir | 31613 | 2004 |
| fulvestrant | 31665 | 2004 |
| gemcitabine | 31666 | 2004 |
| ibandronic acid | 31573 | 2004 |
| ibritumomab tiuxetan | 31604 | 2004 |
| Iloprost | 31558 | 2004 |
| imatinib | 31565 | 2004 |
| imatinib | 31565 | 2004 |
| infliximab | 31623 | 2004 |
| infliximab | 31624 | 2004 |
| infliximab | 31623 | 2004 |
| Laronidase | 31523 | 2004 |
| liposomal cytarabine | 31574 | 2004 |
| mitotane | 31673 | 2004 |
| Mycophenolate sodium | 31573 | 2004 |
| oseltamivir | 31516 | 2004 |
| Pegvisomant | 31497 | 2004 |
| Pioglitazone | 31544 | 2004 |
| pramipexole | 31588 | 2004 |
| Rabeprazole | 31701 | 2004 |
| rivastigmine | 31577 | 2004 |
| ropinirole | 31712 | 2004 |
| Rosiglitazone | 31546 | 2004 |
| salmeterol fluticasone | 31496 | 2004 |
| TachoSil medicated sponge | 31707 | 2004 |
| tenofovir | 31793 | 2004 |
| teriparatide | 31532 | 2004 |
| topiramate | 31693 | 2004 |
| adalimumab | 32293 | 2005 |
| alendronate | 32236 | 2005 |
| Anagrelide | 31926 | 2005 |

| | | |
|---|---|---|
| Aprepitant | 32229 | 2005 |
| azelaic acid | 31882 | 2005 |
| bevacizumab | 31832 | 2005 |
| bivalirudin | 31807 | 2005 |
| Candesartan cilexetil | 322368 | 2005 |
| capecitabine | 31794 | 2005 |
| cetuximab | 31761 | 2005 |
| cinacalcet | 32205 | 2005 |
| levodopa/carbidopa | 32603 | 2005 |
| Darbepoetin alfa | 31725 | 2005 |
| docetaxel | 31922 | 2005 |
| efalizumab | 31789 | 2005 |
| erdosteine | 32038 | 2005 |
| esomeprazole | 31840 | 2005 |
| etanercept | 31779 | 2005 |
| fludarabine | 32288 | 2005 |
| fondaparinux | 32191 | 2005 |
| fondaparinux | 32192 | 2005 |
| galantamine | 31914 | 2005 |
| imiquimod | 31795 | 2005 |
| imiquimod | 31777 | 2005 |
| Insulin detemir | 31784 | 2005 |
| insulin glulisine | 31808 | 2005 |
| letrozole | 32287 | 2005 |
| letrozole | 31762 | 2005 |
| micronised progesterone | 32411 | 2005 |
| Modafinil | 31902 | 2005 |
| Montelukast | 31920 | 2005 |
| mycophenolate mofetil | 31834 | 2005 |
| Nicotinic acid | 31820 | 2005 |
| Olopatadine | 32392 | 2005 |
| oxaliplatin | 32312 | 2005 |
| Oxycodone | 32209 | 2005 |
| Oxycodone | 32210 | 2005 |
| peginterferon alfa-2a | 31910 | 2005 |
| pemetrexed | 31778 | 2005 |
| pregabalin | 31785 | 2005 |
| pregabalin | 31785 | 2005 |
| rituximab | 31899 | 2005 |
| rosuvastatin | 31730 | 2005 |
| rosuvastatin | 32279 | 2005 |
| sirolimus | 31913 | 2005 |
| Strontium ranelate | 31773 | 2005 |
| trastuzumab | 32203 | 2005 |
| Vinorelbine | 32384 | 2005 |
| Vinorelbine | 32283 | 2005 |
| zoledronic acid | 32221 | 2005 |
| Zonisamide | 32329 | 2005 |
| adalimumab | 2893 | 2006 |
| Aprepitant | 3093 | 2006 |
| bortezomib | 32576 | 2006 |
| capecitabine | 32591 | 2006 |
| Carmustine implant | 32363 | 2006 |
| cetuximab | 3135 | 2006 |
| daptomycin | 2927 | 2006 |
| deferasirox | 3381 | 2006 |
| dibotermin alfa | 2410 | 2006 |
| dipyridamole | 32797 | 2006 |
| docetaxel | 32531 | 2006 |
| erlotinib | 2253 | 2006 |

| | | |
|---|---|---|
| exemestane | 2878 | 2006 |
| exenatide | 4012 | 2006 |
| frovatriptan | 3460 | 2006 |
| ibandronic acid | 3518 | 2006 |
| infliximab | 32588 | 2006 |
| infliximab | 32588 | 2006 |
| infliximab | 32588 | 2006 |
| Interferon beta-1b | 3083 | 2006 |
| Ivabradine | 3467 | 2006 |
| lanthanum carbonate | 2164 | 2006 |
| Losartan | 32650 | 2006 |
| Metformin hydrochloride | 32503 | 2006 |
| nebivolol | 2925 | 2006 |
| nebivolol | 2321 | 2006 |
| omalizumab | 2170 | 2006 |
| paclitaxel | 32529 | 2006 |
| parathyroid hormone | 3683 | 2006 |
| pegaptanib | 3148 | 2006 |
| pegaptanib | 3148 | 2006 |
| peginterferon alfa-2a | 32520 | 2006 |
| pemetrexed | 31778 | 2006 |
| posaconazole | 32611 | 2006 |
| pravastatin | 32553 | 2006 |
| pravastatin | 32554 | 2006 |
| pravastatin | 32555 | 2006 |
| raloxifene | 2594 | 2006 |
| rasagiline | 32476 | 2006 |
| Risperidone | 32570 | 2006 |
| rituximab | 3446 | 2006 |
| rituximab | 3723 | 2006 |
| rivastigmine | 2879 | 2006 |
| Rosiglitazone | 32598 | 2006 |
| sildenafil | 2255 | 2006 |
| simvastatin | 32466 | 2006 |
| simvastatin | 32620 | 2006 |
| sodium oxybate | 2782 | 2006 |
| Solifenacin | 1884 | 2006 |
| sorafenib | 2905 | 2006 |
| Strontium ranelate | 3304 | 2006 |
| sunitinib | 3144 | 2006 |
| sunitinib | 3144 | 2006 |
| TachoSil medicated sponge | 1043 | 2006 |
| temozolomide | 32614 | 2006 |
| teriparatide | 3303 | 2006 |
| testosterone undecanoate | 3318 | 2006 |
| tiagabine | 32501 | 2006 |
| tigecycline | 3108 | 2006 |
| Tipranavir | 2482 | 2006 |
| Topotecan | 32608 | 2006 |
| trastuzumab | 3054 | 2006 |
| Travoprost | 3731 | 2006 |
| vigabatrin | 2522 | 2006 |
| abatacept | 5592 | 2007 |
| adalimumab | 4863 | 2007 |
| adefovir dipivoxil | 4155 | 2007 |
| alendronate | 4223 | 2007 |
| betaine anhydrous | 5071 | 2007 |
| bortezomib | 4139 | 2007 |
| Budesonide/eformoterol | 4315 | 2007 |
| Calcipotriol and betamethasone | 4258 | 2007 |

| | | |
|---|---|---|
| clopidogrel | 4159 | 2007 |
| levodopa/carbidopa | 4717 | 2007 |
| darunavir | 4442 | 2007 |
| dasatinib | 4070 | 2007 |
| dexrazoxane | 4069 | 2007 |
| docetaxel | 4136 | 2007 |
| donepezil | 3132 | 2007 |
| drotrecogin alfa | 824 | 2007 |
| duloxetine | 2319 | 2007 |
| esomeprazole | 4795 | 2007 |
| esomeprazole | 3729 | 2007 |
| fondaparinux | 4111 | 2007 |
| fondaparinux | 5091 | 2007 |
| gabapentin | 3850 | 2007 |
| galantamine | 3146 | 2007 |
| Glyceryl trinitrate | 4195 | 2007 |
| idursulfase | 4169 | 2007 |
| infliximab | 4627 | 2007 |
| lamotrigine | 4625 | 2007 |
| Latanoprost | 2984 | 2007 |
| lenalidomide | 4856 | 2007 |
| levetiracetam | 3913 | 2007 |
| memantine | 4352 | 2007 |
| methadone | 5198 | 2007 |
| Methoxy polyethylene glycol-epoetin beta | 5111 | 2007 |
| naltrexone | 4027 | 2007 |
| natalizumab | 3657 | 2007 |
| paliperidone | 5168 | 2007 |
| Palonosetron | 3936 | 2007 |
| posaconazole | 4628 | 2007 |
| posaconazole | 4628 | 2007 |
| pramipexole | 3918 | 2007 |
| pregabalin | 4024 | 2007 |
| risedronate | 4544 | 2007 |
| sitaxentan | 4468 | 2007 |
| sodium oxybate | 4626 | 2007 |
| sunitinib | 4512 | 2007 |
| telbivudine | 4976 | 2007 |
| topiramate | 4325 | 2007 |
| zoledronic acid | 5982 | 2007 |
| zopiclone | 4439 | 2007 |
| adalimumab | 5381 | 2008 |
| alemtuzumab | 5602 | 2008 |
| aliskiren | 5216 | 2008 |
| ambrisentan | 5603 | 2008 |
| anidulafungin | 5880 | 2008 |
| anidulafungin | 5315 | 2008 |
| aripiprazole | 5027 | 2008 |
| atazanavir | 5432 | 2008 |
| bevacizumab | 5075 | 2008 |
| bevacizumab | 5390 | 2008 |
| botulinum neurotoxin type A | 5377 | 2008 |
| buprenorphine/naloxone | 5344 | 2008 |
| capecitabine | 5238 | 2008 |
| Cilostazol | 5444 | 2008 |
| clobetasol propionate | 5185 | 2008 |
| Clostridium botulinum type A neurotoxin | 5377 | 2008 |
| docetaxel | 5360 | 2008 |
| entecavir | 3487 | 2008 |
| escitalopram | 5257 | 2008 |

| | | |
|---|---|---|
| etanercept | 5041 | 2008 |
| etanercept | 5041 | 2008 |
| etanercept | 5041 | 2008 |
| etanercept | 5041 | 2008 |
| etanercept | 5041 | 2008 |
| fluticasone furoate | 5384 | 2008 |
| fondaparinux | 5418 | 2008 |
| fondaparinux | 5405 | 2008 |
| lapatinib | 5358 | 2008 |
| lidocaine | 5354 | 2008 |
| maraviroc | 5283 | 2008 |
| mycophenolate mofetil | 5616 | 2008 |
| nilotinib | 5206 | 2008 |
| Olopatadine | 5271 | 2008 |
| oseltamivir | 5375 | 2008 |
| Oxycodone | 5917 | 2008 |
| peginterferon alfa-2a | 5224 | 2008 |
| peginterferon alfa-2b | 4821 | 2008 |
| peginterferon alfa-2b | 5292 | 2008 |
| rasagiline | 5435 | 2008 |
| risedronate | 5491 | 2008 |
| rituximab | 5656 | 2008 |
| rotigotine | 5382 | 2008 |
| sevelamer | 5276 | 2008 |
| sevelamer | 5305 | 2008 |
| sirolimus | 3455 | 2008 |
| sitagliptin | 4513 | 2008 |
| sorafenib | 5225 | 2008 |
| tacrolimus | 5329 | 2008 |
| tacrolimus | 4886 | 2008 |
| tacrolimus | 5237 | 2008 |
| teriparatide | 5200 | 2008 |
| teriparatide | 5572 | 2008 |
| trabectedin | 5252 | 2008 |
| zanamivir | 4715 | 2008 |
| ziconotide | 5245 | 2008 |
| aliskiren | 6371 | 2009 |
| alitretinion | 6204 | 2009 |
| alteplase | 6722 | 2009 |
| aripiprazole | 5831 | 2009 |
| aripiprazole | 6282 | 2009 |
| atazanavir | 6283 | 2009 |
| bevacizumab | 5479 | 2009 |
| bevacizumab | 6250 | 2009 |
| bivalirudin | 6275 | 2009 |
| bortezomib | 6647 | 2009 |
| capecitabine | 6576 | 2009 |
| Cinacalcet | 6202 | 2009 |
| dabigatran | 5528 | 2009 |
| darunavir | 6571 | 2009 |
| darunavir | 6572 | 2009 |
| darunavir | 6833 | 2009 |
| doripenem | 5849 | 2009 |
| esomeprazole | 6152 | 2009 |
| etravirine (Intelence) | 6000 | 2009 |
| ezetimibe | 6429 | 2009 |
| ezetimibe | 6429 | 2009 |
| ezetimibe | 6429 | 2009 |
| febuxostat | 6315 | 2009 |
| fentanyl | 6829 | 2009 |

| | | |
|---|---|---|
| icatibant | 5904 | 2009 |
| imiquimod | 5614 | 2009 |
| lacosamide | 6048 | 2009 |
| methylnaltrexone bromide | 5881 | 2009 |
| micafungin | 5880 | 2009 |
| Modafinil | 2921 | 2009 |
| nebivolol | 6325 | 2009 |
| nebivolol | 6325 | 2009 |
| peginterferon alfa-2b | 5474 | 2009 |
| pegylated liposomal doxorubicin | 6273 | 2009 |
| pemetrexed | 5800 | 2009 |
| rivaroxaban | 6017 | 2009 |
| Rosiglitazone | 3161 | 2009 |
| rosuvastatin | 5927 | 2009 |
| rufinamide | 6044 | 2009 |
| salmeterol fluticasone | 5503 | 2009 |
| sugammadex | 6014 | 2009 |
| Temoporfin | 5910 | 2009 |
| tenofovir | 6085 | 2009 |
| thalidomide | 5573 | 2009 |
| Topotecan | 5514 | 2009 |
| vildagliptin | 5731 | 2009 |
| Vinorelbine | 6288 | 2009 |
| Vinorelbine | 6287 | 2009 |
| zoledronic acid | 6147 | 2009 |

**HAS Dataset: Missing Data**

Between January 2004-June 2009, the Comittee de Transparence part of the Haute Autorite de Sante (HAS) reviewed 315 submissions and made specific recommendations as to the incremental medical value of the new therapies under review.

A data set of information pertaining to HAS appraisals was created collecting information on variables relating to (i) the clinical and economic characteristics of the technology under appraisal, as well as information on (ii) the process used to come to a decision, and (iii) the socio-economic context in which these decisions were made. In order to prepare the data set for analysis, understanding the data set is important. Here, the aim is to characterise the presence of missing data across variables and decisions to inform the need for imputing missing data.

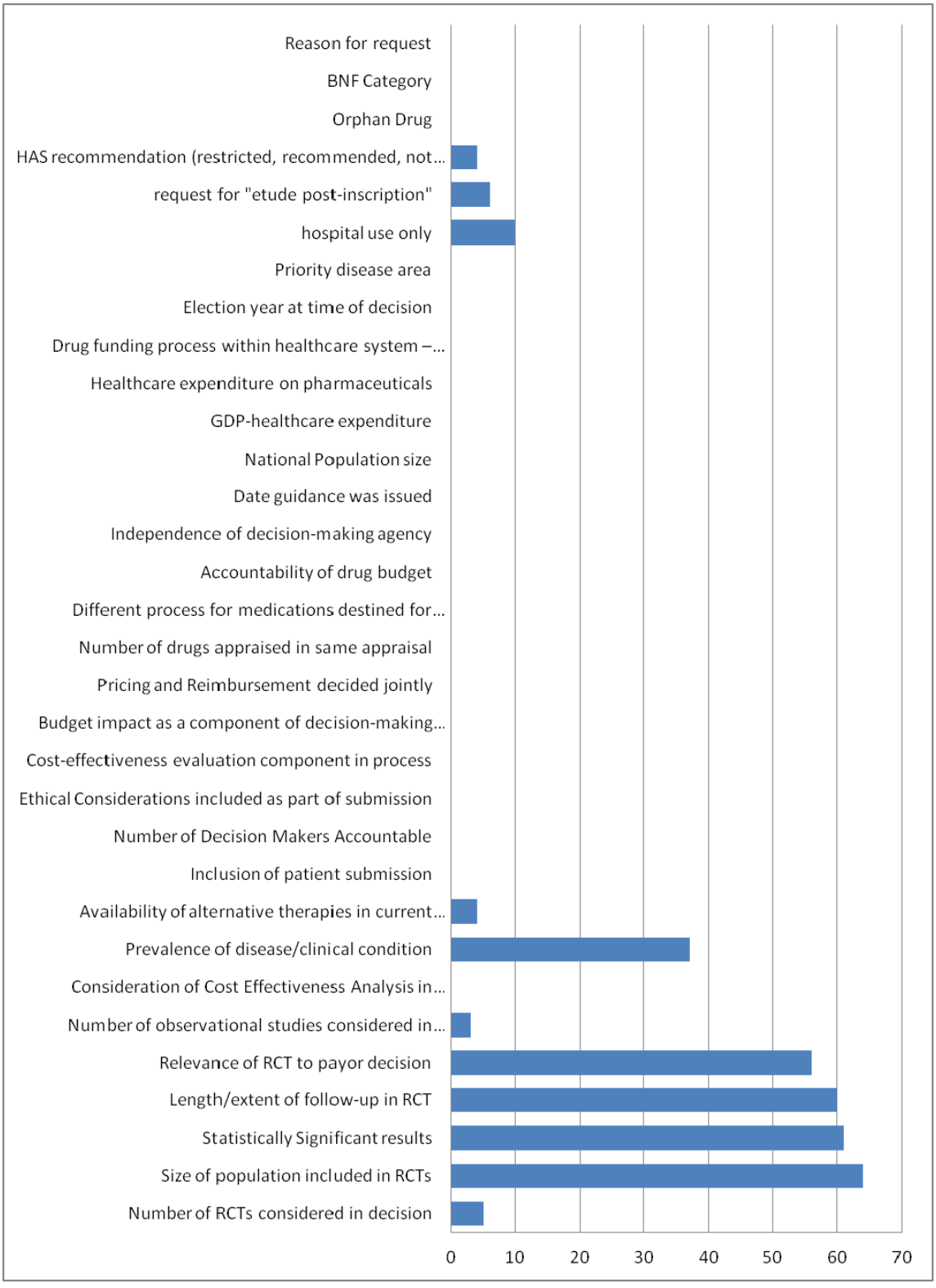
*Distribution of Missing Data within HAS data set*

The outcome considered is the ASMR rating given by the HAS committee, which ranges from 1 (high incremental beneft) to 5 (no incremental benefit). In the total HAS sample, there are 5% of missing entries.

The distribution of missing data **across each variable** was examined. The total number of observations per variable is 315. The variables with the highest number of missing information are those related to the percentage reimbursement (‘taux de remboursement) and clinical package. Note that the HAS does not include as part of the review process any cost components.

The extent of missing data was also examined **across appraisals**. The level of missing variables per appraisal ranged from 0 – 8. The average number of ‘missing entries per appraisal was less than 1.

Figure D.1. HAS Distribution of missing data by variable (n=315)



**HAS Dataset – Testing of Normality Assumption**

**Method**

To determine the relevant statistical tests to use in assessing the significance of differences observed between means, it was necessary to assess whether the normality assumption was valid for the variables under consideration. This would then determine the use of parametric or non-parametric tests. For all variables, the sample is >30. It has been suggested that when analysing sample sizes of >30, even when the normality assumption is violated, parametric tests may still be performed (Pallant 2007, SPSS Survival Manual, 3rd edn, Maidenhead, OUP/McGraw-Hill). Prior to making a decision on which variables to apply parametric or non-parametric tests, the distribution for each variable will be further examined.

To test the normality assumption the following was performed for each non-categorical variable:

- Skeweness was calculated
- Kurtosis was calculated
- The standardized normal probability plot (P-P plot) was graphed

The graph and calculations are presented for each non-categorical variable and an assessment is made of whether the normality assumption is met or not. A variable will be considered to meet the normal distribution assumption if

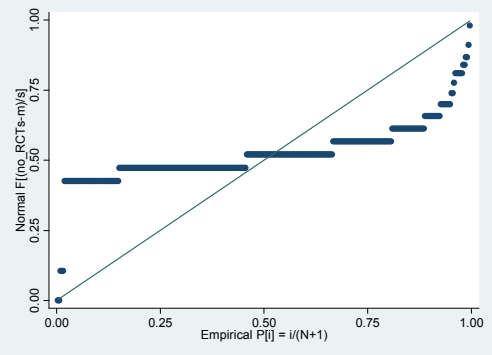
- skewness is within the range ±1
- Kurtosis value is within range ±3
- P-P plot shows distribution of dataset is approximately linear

In those circumstances where there is inconsistency between the three tests, if 2 of the tests suggest normal distribution, then it will be considered as such (but then tested using both parametric and non-parametric tests in sensitivity analyses).

**Number of RCTs considered in appraisal (No_RCT)**

Skewness 2.583494

Kurtosis 13.64883

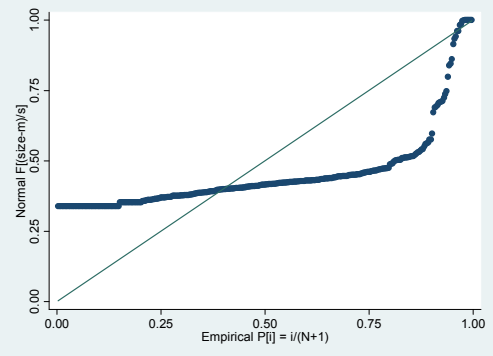



This variable is **not** normally distributed.

**Mean sample size of RCTs considered in appraisal (RCTsize)**

Skewness 5.551863

Kurtosis 39.18887




This variable is **not** normally distributed.

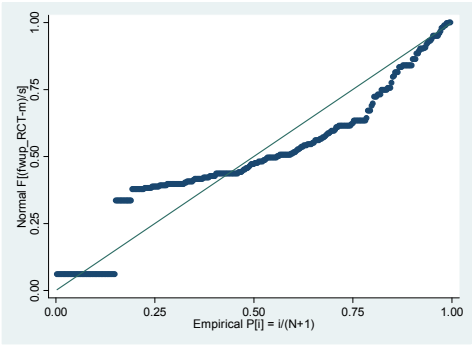
**Superiority demonstrated**

Normality test not performed as this is a categorical variable.

**Duration of RCT (RCTfwup)**

Skewness 3.850601

Kurtosis 25.89232




This variable is **not** normally distributed.

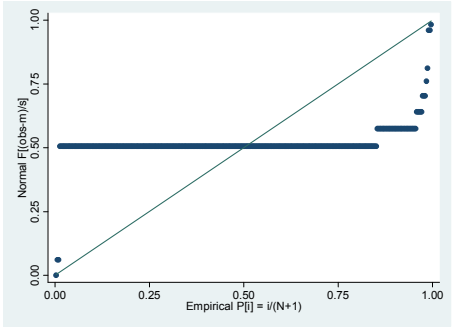
**Active-Comp**

Categorical variable.

**Observational Studies**

Skewness 7.092638

Kurtosis 60.95931

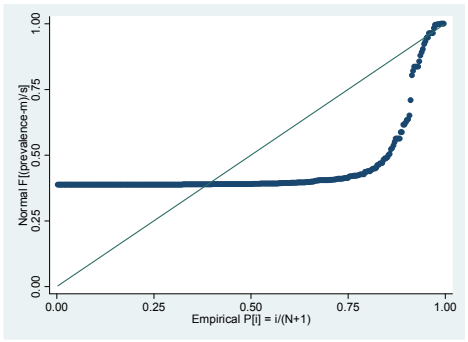



This variable is **not** normally distributed.

**Prevalence**

Skewness 5.9439

Kurtosis 44.58945




This variable is **not** normally distributed.

**Alternative Available**

Normality test not performed as this is a categorical variable.

**Ethics**

Normality test not performed as this is a categorical variable.

**Cost-Effectiveness part of Process**

Normality test not performed as this is a categorical variable.

**BIM part of Process**

Normality test not performed as this is a categorical variable.

**Joint reimbursement & pricing decision**

Normality test not performed as this is a categorical variable.

**Accountability for drug budget**

Normality test not performed as this is a categorical variable.

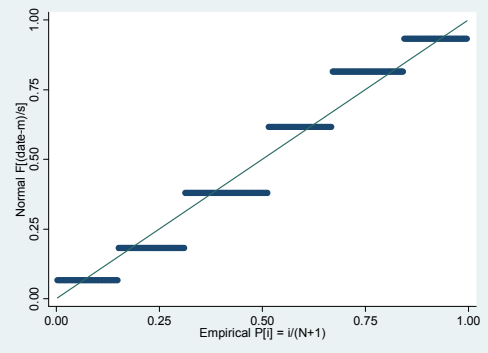
**Independence of HTA agency from MoH**

Normality test not performed as this is a categorical variable.

**Year of appraisal**

Skewness .0097271

Kurtosis 1.799967

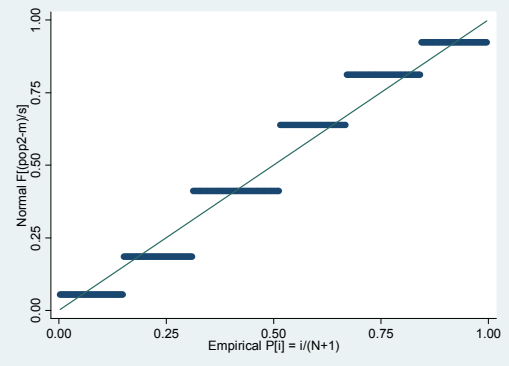



Perhaps this is actually a categorical variable?

**Population under HTA remit**

Skewness -.1608239

Kurtosis 1.858135




Perhaps this is actually a categorical variable?

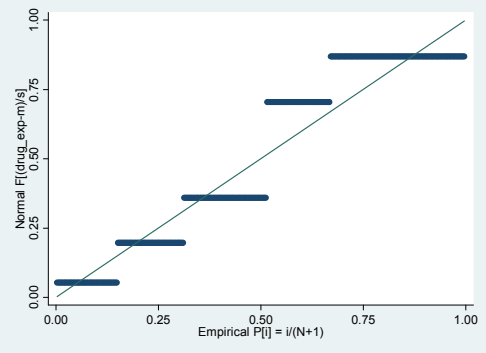
**% of GDP expenditure on health**

Same percentage GDP level between 2004-2009. Therefore, no variation in the distribution across groups.

**Drug expenditure per person**

Skewness -.257072

Kurtosis 1.663579




**Funding mechanism – centralised or decentralised**

Normality test not performed as this is a categorical variable.

**Election year**

Normality test not performed as this is a categorical variable.

**Priority Disease Area**

Normality test not performed as this is a categorical variable.

**BNF Category**

Normality test not performed as this is a categorical variable.

## HAS dataset – descriptive statistics

| | Chi2 | ANOVA | T-Test (Rec vs Res) | T-Test (Res vs NR) | Kruskal-Wallis |
|---|---|---|---|---|---|
| Number of RCTs considered in decision | | 0.013 | 0.0415 | 0.5424 | 0.3375 |
| Size of population included in RCTs | | 0.8813 | 0.5764 | 0.9895 | 0.2038 |
| Statistically Significant results - yes (1) | 0.0001 | | | | |
| no (0) | - | - | - | - | - |
| inconsistent (2) | - | - | - | - | - |
| Length/extent of follow-up in RCT | | 0.1757 | 0.3324 | 0.3474 | 0.0013 |
| Relevance of RCT to payor decision | | 0.074 | 0.0716 | 0.1067 | 0.0137 |
| Number of observational studies considered in guidance | | 0.2197 | 0.487 | 0.2443 | 0.9466 |
| Consideration of CUA in guidance | - | - | - | - | - |
| Prevalence of disease | | 0.0156 | 0.0213 | 0.0033 | 0.0001 |
| Availability of alternative therapies in current treatment setting. | 0.014 | | | | |
| Inclusion of patient submission | - | - | - | - | - |
| Number of Decision Makers Accountable | - | - | - | - | - |
| Ethical Considerations included as part of submission | - | - | - | - | - |
| Cost-effectiveness evaluation component in process | - | - | - | - | - |
| Budget impact as a component of decision-making process | - | - | - | - | - |
| Pricing and Reimbursement decided jointly | - | - | - | - | - |
| Accountability of drug budget | - | - | - | - | - |
| Independence of decision-making | - | - | - | - | - |
| Date guidance was issued | | 0.009 | 0.1108 | 0.0214 | 0.0017 |
| Population size – Agency coverage | | 0.009 | 0.1384 | 0.0185 | 0.0017 |
| GDP-healthcare expenditure | | n/a | n/a | n/a | n/a |
| Healthcare expenditure on pharmaceuticals | | 0.0043 | 0.1865 | 0.0101 | 0.0019 |
| Drug funding process within healthcare system – whether centralized or decentralised | - | - | - | - | - |
| Election year at time of decision | 0.039 | | | | |
| Priority disease area | | | | | |
| Orphan Designated | 0.0001 | | | | |
| Taux de remboursement | | 0.0646 | 0.9737 | 0.0082 | 0.0864 |
| HAS recommendation | 0.009 | | | | |
| recommended | - | - | - | - | - |
| restricted | - | - | - | - | - |
| not recommended | - | - | - | - | - |
| Hospital use only | 0.0001 | | | | |
| Request for post-marketing study | 0.004 | | | | |
| Reason for request | 0.094 | | | | |
| Inscription Sécurité Sociale et Collectivités | - | - | - | - | - |
| Réévaluation | - | - | - | - | - |
| extension d'indication | - | - | - | - | - |
| renouvellement d'inscription | - | - | - | - | - |
| Redéfinition du périmètre des indications remboursables | - | - | - | - | - |
| BNF Category | 0.0001 | | | | |

## HAS preliminary multivariate analysis model

**Table 7.1 Multivariate analysis of HAS coverage decisions 2004-2009: preliminary model (n=315)**

| ASMR III-IV | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Number of RCTs | -0.2843941 | 0.044 | -0.5611727 | -0.00762 |
| Clinical superiority demonstrated in RCT | -3.436836 | 0.146 | -8.070138 | 1.196467 |
| Lack of clinical superiority in RCT | -3.545748 | 0.146 | -8.331433 | 1.239936 |
| Inconsistent demonstration of clinical superiority in RCT | -2.217476 | 0.334 | -6.718472 | 2.283521 |
| RCT duration of follow-up | -0.0033383 | 0.308 | -0.0097539 | 0.003077 |
| Use of active comparator in RCT | 0.7665676 | 0.189 | -0.3780955 | 1.911231 |
| Disease prevalence | 5.22E-07 | 0.182 | -2.45E-07 | 1.29E-06 |
| Presence of alternative therapy | 0.6985936 | 0.04 | 0.0317558 | 1.365431 |
| Year guidance was issued | 1.109892 | 0.136 | -0.3503391 | 2.570123 |
| Orphan designation status | -0.5771616 | 0.359 | -1.809676 | 0.655353 |
| Malignant disease and immunosuppression | -0.1716978 | 0.826 | -1.700147 | 1.356752 |
| Central nervous system | 4.623967 | 0.015 | 0.9009441 | 8.346989 |
| Infections | 2.744761 | 0.039 | 0.1401706 | 5.349351 |
| Musculoskeletal and joint diseases | -1.2293 | 0.064 | -2.528295 | 0.069695 |
| Respiratory system | 23.22601 | 0.000 | 21.32679 | 25.12522 |
| Skin | 21.50354 | 0.000 | 20.19059 | 22.81649 |
| Hospital use only | -0.2385087 | 0.745 | -1.675585 | 1.198568 |
| Healthcare expenditure on pharmaceuticals | -0.07369 | 0.172 | -0.1793296 | 0.03195 |
| Reimbursement level obtained | -0.2492197 | 0.917 | -4.920654 | 4.422215 |
| Missing data on numbers of RCTs considered | 20.86546 | . | . | . |
| Missing data on duration of RCTs | -0.0353642 | 0.984 | -3.551916 | 3.481188 |
| Missing data on active comparators used in RCT | -4.462134 | 0.071 | -9.303942 | 0.379675 |
| Missing data on prevalence of disease | -2.325411 | 0.005 | -3.964093 | -0.68673 |
| Missing data on level of reimbursement | -0.337959 | 0.463 | -1.240771 | 0.564853 |
| Missing data on hospital use | 0.1355593 | 0.903 | -2.036081 | 2.3072 |
| Constant | -2184.194 | 0.136 | -5059.288 | 690.8998 |
| ASMR V | Log Odds | P value | 95% Conf. Interval | |
| Number of RCTs | -0.3918877 | 0.007 | -0.6766837 | -0.10709 |
| Clinical superiority demonstrated in RCT | 14.13418 | 0.992 | -2883.015 | 2911.283 |
| Lack of clinical superiority in RCT | 15.67453 | 0.992 | -2881.435 | 2912.784 |
| Inconsistent demonstration of clinical superiority in RCT | 16.1639 | 0.991 | -2881.056 | 2913.383 |
| RCT duration of follow-up | -0.0042145 | 0.229 | -0.0110837 | 0.002655 |
| Use of active comparator in RCT | 1.101055 | 0.074 | -0.1066088 | 2.308718 |
| Disease prevalence | 9.50E-07 | 0.017 | 1.70E-07 | 1.73E-06 |
| Presence of alternative therapy | 0.5738384 | 0.353 | -0.6380268 | 1.785704 |
| Year guidance was issued | 0.6479199 | 0.388 | -0.8234194 | 2.119259 |
| Orphan designation status | -3.089502 | 0.001 | -4.989264 | -1.18974 |

| | | | | |
|---|---|---|---|---|
| Malignant disease and immunosuppression | 0.2362465 | 0.785 | -1.459433 | 1.931926 |
| Central nervous system | 4.959432 | 0.01 | 1.207159 | 8.711704 |
| Infections | 2.691694 | 0.045 | 0.0651163 | 5.318272 |
| Musculoskeletal and joint diseases | -2.397214 | 0.003 | -3.987509 | -0.80692 |
| Respiratory system | 23.52017 | . | . | . |
| Skin | 21.35004 | . | . | . |
| Hospital use only | -0.1694243 | 0.826 | -1.679657 | 1.340808 |
| Healthcare expenditure on pharmaceuticals | -0.0245484 | 0.654 | -0.1317318 | 0.082635 |
| Reimbursement level obtained | -2.667338 | 0.281 | -7.517837 | 2.18316 |
| Missing data on numbers of RCTs considered | 21.26527 | 0.000 | 18.82152 | 23.70903 |
| Missing data on duration of RCTs | -0.9071335 | 0.686 | -5.310142 | 3.495875 |
| Missing data on active comparators used in RCT | 14.28305 | 0.992 | -2883.184 | 2911.75 |
| Missing data on prevalence of disease | -2.916096 | 0.001 | -4.578727 | -1.25346 |
| Missing data on level of reimbursement | -0.1922116 | 0.711 | -1.207439 | 0.823016 |
| Missing data on hospital use | -0.5220553 | 0.693 | -3.112505 | 2.068394 |
| Constant | -1298.865 | . | . | . |

Note: Technologies with ASMR I or II are the reference case.

## E. Summary of Interviews with representatives of NICE, SMC, and CFH

| HTA Body | **NICE** |
|---|---|
| Date | 2nd of February 2011 |
| Attendees | Professor Peter Littlejohns<br>Clinical and Public Health Director<br>National Institute for Health and Clinical Excellence<br>MidCity Place \| 71 High Holborn \| London WC1V 6NA \| United Kingdom<br>Tel: 44 (0)20 7045 2091 fax: 44 (0)845 003 7784 |
| Objectives | The aim of this interaction was to ascertain if the NICE characteristics were accurately captured in the sample used for analysis, if the approach to analysis was clear and in particular the reaction to the model results and potential for suggestions or additional analyses. |
| Discussion Summary | Sample Characteristics<br>o Prof. Littlejohns confirmed that a three-category outcome variable is a valid option to capturing the decision-making options available to the NICE. It was recognised that both three-category and binary outcome variables had also been used in published assessments of NICE decision-making.<br>o The reported proportion of recommendations, restrictions and non-recommendations observed within the sample of analysis was felt to differ from other publically available statistics on the same.<br>o Professor Littlejohns provided information on the distribution of coverage decisions made by NICE in 2000-2010 for all technology types (including medical devices) and patient populations (including paediatrics).<br>o The proportions differed from those observed within the analysis. It was discussed that the potential sources for these differences could be due to a different time horizon and technologies included. This dataset incorporated decisions made between January 2004 and June 2009, and included pharmaceuticals indicated for adult populations.<br>o It was felt to be a plausible explanation for the differences observed.<br>o The explanatory variables extracted for analysis were felt to be appropriate and reflect various aspects of the NICE decision-making process.<br>Model results<br>• A unit increase in the ICER was found to increase the odds of restriction or non-recommendation and in the discussion this was seen as reflecting NICE decision-making and coherent with previous published research<br>• The importance of clinical criteria shown in the model |

| | |
|---|---|
| | was also considered to be coherent with NICE's decision-making process<br>• An increase in the number of technologies appraised within the same appraisal was found to increase the odds of restriction. While this had had not previously been studied in other published literature, Professor Littlejohns offered the interpretation that it could be plausible to assume that the appraisal of multiple technologies simultaneously may lead to a situation in which a single technology is recommended, and the rest are restricted or not recommended.<br>• Professor Littlejohns stressed the importance of ascertaining the distribution of NICE decisions across the three outcome categories to ensure the validity of the results, but overall was able to state that the explanatory variables identified as significant in the model could plausibly reflect NICE decision-making. |

| HTA Body | **SMC** |
|---|---|
| Date | 20th December 2010 |
| Attendees | Dr Andrew Walker,<br>Member of the SMC NDC Senior Lecturer in Health Economics, University of Glasgow |
| Objectives | The aim of this interaction was to ascertain if the SMC characteristics were accurately captured in the sample used for analysis, if the approach to analysis was clear and in particular the reaction to the model results and potential for suggestions or additional analyses. |
| Discussion Summary | Role of SMC<br>• Dr. Walker commented that SMC does not assess the notion of affordability, unlike NICE.<br>• Highlighted that the SMC issues decisions that either *accept* or reject technologies for use, and this subtlety should be reflected in how SMC outcomes are described<br>• Dr. Walker confirmed that a three-category outcome variable accurately reflected the decision-making options available to the SMC. It was noted that the proportion of recommendation was perhaps lower than what would be anticipated, although it was suggested that this is likely explained by the time horizon used in the analysis (2004-2009), and to the fact that the analysis was restricted to full submissions and resubmissions (i.e. excluded abbreviated submissions).<br>• The explanatory variables extracted for analysis were felt to be appropriate and reflect various aspects of the SMC decision-making process.<br>Model results<br>• Draft multivariate analysis of SMC decision-making was shared to gather feedback on the internal validity of the model.<br>• An increasing trial duration was found to decrease the log odds of a restriction and non-recommendation, and this effect was found to be plausible within the SMC decision-making process which places significant emphasis on the clinical evidence and quality of that evidence<br>• An increasing ICER decreased the odds of recommendation and this was felt to reflect the SMC's objective of determining and accepting technologies that demonstrated value-for-money.<br>• The role of the budgetary impact was discussed at length. The draft model suggested that increasing budgetary impact was associated with an increased odds of recommendation. The first comment was that this confirmed the results of descriptive analysis, suggesting |

| | |
|---|---|
| | that SMC is not driven by budgetary impact concerns. This would also reflect the notion that SMC does not concern itself with the notion of affordability in the same way that NICE would. Dr. Walker suggested that the direction of the effect associated with budget impact in the model may actually be a surrogate or proxy for other variables. For example, it was plausible that technologies with orphan designation or with no alternative therapies could be associated with higher budgetary impact, but were also technologies more likely to be supported by weaker clinical evidence (non-comparative or placebo controlled trials, small sample size, short duration). It was also suggested that technologies with weaker value-for-money profiles may argue for budgetary neutrality as an argument for recommendation, and that the model reflected this strategy. It was recommended by Dr. Walker to further examine the role of the budget impact variable in the model and consider alternative variables to explain this effect.<br>• The impact of an indication for the treatment of skin diseases was also discussed. It was not clear why such technologies, which represent a small proportion of technologies appraised by the SMC, should increase the odds of recommendation. It was suggested by Dr. Walker to further explore and confirm the effect of this variable in the model.<br>• The impact of an indication for infectious diseases was also discussed, particularly the reason for its diverging effect: an infectious diseases indication appeared to decrease the odds of a non-recommendation but increase the odds of a restriction. It was proposed that a plausible reason for this could be due to the fact that antibiotics are included as part of the infectious diseases group, and that while these technologies may not always be supported by high quality evidence there is reluctance to not-recommend antibiotics because of the problems encountered with the development of resistance to antibiotics in clinical practice.<br>• The effect of patient submission in the model decreased the odds of recommendation. Dr. Walker suggested that this may reflect the fact that patient submissions tend to be made for diseases for which there is concern that a non-recommendation is plausible, or for rare and severe diseases. It should not be interpreted as reflecting the fact that patient evidence is not considered or that it impacts negatively on the decision outcome, but rather that it is a proxy for the characteristics of the technology being appraised. It was suggested that further exploration should be conducted to understand the characteristics of those technologies with and without patient submissions.<br>• An overall comment made by Dr. Walker was the need to |

| | |
|---|---|
| | comment on those variables that were not found to be significant, as this is potentially as important as those that were found to be significant in their effect on SMC decisions<br>• Overall, Dr. Walker commented that the variables found to be significant in the model reflected SMC decision-making, with the exception of technologies for skin diseases which needed further confirmation. In addition, he stressed the need to re-examine in more depth the mechanisms through which the budget impact and patient submissions were operating in the model. |

| HTA Body | **CFH of CVZ** |
|---|---|
| Date | 6th of January 2011 |
| Attendees | Dr. Martin van der Graaff, Secretary medicines evaluation committee (CFH) Dr. Wim Goettsch<br>Sarah Kleijnen (M. Sc. ) Project coordinator EUnetHTA WP5<br><br>College voor Zorgverzekeringen (CVZ)<br>Health Care Insurance Board<br>Eekholt 4 1112 XH Diemen<br>Postbus 320 1110 AH Diemen |
| Objectives | The aim of this interaction was to ascertain if the CFH characteristics were accurately captured in the sample used for analysis, if the approach to analysis was clear and in particular the reaction to the model results and potential for suggestions or additional analyses. |
| Discussion Summary | CVZ characteristics<br><br>• The interpretation of the analysis should take into account the fact that CVZ appraisals are not made in a 'static' context - in other words, there are changers over time in the processes that guide the appraisals, and in the outlook of the agency overtime.<br>• Should refer to the CVZ not the CFH<br>• Orphan drug list does not exist per se: orphan technologies are either included like other technologies in the GVS (usually Iist 1B) or on the expensive drug list.<br>• Not all expensive drugs are on the expensive drug list.<br><br>Model variables<br><br>• the categorisation of CVZ coverage decisions into 'recommended', 'restricted', 'not recommended' categories was found to make sense, although it was recognised that such categorisation to some degree over simplifies the reimbursement options in the Dutch system.<br>• Expensive drugs are appraised indication by indication, and inclusion on the expensive drug list is specific to the appraised indication. Could therefore consider inclusion on expensive drug list as a form of restriction<br>• Would be useful to capture severity of disease as a factor, and also end of life technologies. Currently, only disease categories are captured<br>• CVZ decision-making is driven by the strength and the robustness of the scientific evidence supporting the technology. |

| | |
|---|---|
| | Cost effectiveness<br><br>• The use of cost effectiveness in the appraisal process:<br>  o officially part of the process as of 2006, and was used experimentally until 2007-08 when CE criteria became more strongly used in decision-making. Note that if a technology brings no added value, then no Cost-Effectiveness analysis is referred to in the report, even though it may have been submitted by the manufacturer.<br>• In the CVZ appraisal process, used of cost effectiveness results is very different from NICE/SMC; threshold is not utilised yet. Currently, the robustness/quality of the analysis drives the decision. In the future also the outcome of a cost-effectiveness analysis in Euro/QALY may play a role in the decision.<br><br>Model results<br><br>o in general, model results appear to fit with expectations. For example: the presence of an active comparator and the demonstration of superiority in a clinical trial decrease the odds of a restriction/ non recommendation. This was considered to make sense by the participants.<br>o The fact that cancer therapies were associated with an increased odds of a recommendation also made sense, as did the fact that technologies for Cardio-vascular, musculoskeltal diseases decreased the odds of recommendation.<br>o Increase in budget impact increased the odds of a restriction or non-recommendation.<br>o The impact of patient submission was discussed. Appraisals for the 1b or expensive drug list frequently include the request for feedback from patient groups - however, these are not often reported in the final CVZ reports. This strongly suggests that the proportion of technologies with patient submissions identified in the publicly available data is under-represented. |

# F. Imputation techniques – summary of imputation techniques tested using SMC dataset

**Objective**

To maximize the number of observations and sample size for both the individual and pooled analyses, imputation techniques were used to estimate entries for those observations that were affected by 'not applicable' or 'not reported' data. In order to determine the method most adapted to the analyses, several approaches to imputation were tested including imputation by replacing missing values with the overall mean of the variable, generating regression estimates of the missing value, and multiple imputation techniques that take random variation into account. This exercise was performed on one of the HTA datasets (SMC dataset) to test which imputation method would be most useful to extrapolate to the remaining HTA multivariate analyses.

**Analyses**

The results of this exercise are summarised in the tables provided below:

- Table 1 provides the multinomial logistic regression results in which incomplete observations were imputed using mean imputation. In this technique, the mean value of the variable is used to fill incomplete observations.
- Table 2 provides the multinomial logistic regression results in which incomplete observations were imputed using regression driven imputation. In this technique, regression from other available variables is used to generate a value to fill incomplete entries. To generate these values the 'impute' command from Stata was used.
- Table 3 provides the multinomial logistic regression results in which incomplete observations were imputed using multiple imputation techniques. In this technique, the statistical model to generate the imputed values is based on multiple imputed datasets are created each of which contain different imputed values (UCLA 2011). This technique can also take into consideration whether the variable is categorical, and was therefore felt to be relevant for this analysis. In this particular test case, only one imputed dataset was generated.

**Table 1. SMC multinomial logistic regression results <u>With mean imputation</u>**

Multinomial logistic regression — Number of obs = 288; LR chi2(14) = 64.99; Prob > chi2 = 0.0000; Log likelihood = -266.75563; Pseudo R2 = 0.1086

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **restricted** | | | | | | |
| RCTsize | .000034 | .0000643 | 0.53 | 0.596 | -.0000919 | .00016 |
| RCTfwup | -.0030883 | .0029582 | -1.04 | 0.296 | -.0088862 | .0027096 |
| ICER | .0000353 | .0000133 | 2.65 | 0.008 | 9.19e-06 | .0000613 |
| BNF7 | -.4321934 | .4895508 | -0.88 | 0.377 | -1.391695 | .5273085 |
| BNF13 | -.3657278 | .7352662 | -0.50 | 0.619 | -1.806823 | 1.075368 |
| prevalence | -.0000124 | 8.44e-06 | -1.47 | 0.141 | -.000029 | 4.12e-06 |
| alternative | -.6010864 | .6107646 | -0.98 | 0.325 | -1.798163 | .5959902 |
| _cons | .8277638 | .6830546 | 1.21 | 0.226 | -.5109986 | 2.166526 |
| **not recomm~d** | | | | | | |
| RCTsize | -.000307 | .0001717 | -1.79 | 0.074 | -.0006435 | .0000295 |
| RCTfwup | -.0087137 | .0036469 | -2.39 | 0.017 | -.0158615 | -.001566 |
| ICER | .0000382 | .0000133 | 2.88 | 0.004 | .0000122 | .0000643 |
| BNF7 | -1.965731 | .5906051 | -3.33 | 0.001 | -3.123296 | -.8081663 |
| BNF13 | -1.818445 | .9247152 | -1.97 | 0.049 | -3.630854 | -.006037 |
| prevalence | -.0000209 | 9.76e-06 | -2.14 | 0.033 | -.00004 | -1.74e-06 |
| alternative | -1.042411 | .6076999 | -1.72 | 0.086 | -2.233481 | .1486584 |
| _cons | 2.099526 | .6853151 | 3.06 | 0.002 | .7563334 | 3.442719 |

(decision==recommended is the base outcome)

**Table 2. SMC multinomial logistic regression results <u>With imputation driven by regression from other available variables (i.e. impute command in stata)</u>**

Multinomial logistic regression — Number of obs = 288; LR chi2(14) = 61.98; Prob > chi2 = 0.0000; Log likelihood = -268.25905; Pseudo R2 = 0.1036

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **restricted** | | | | | | |
| IMPRCTsize | .0000563 | .0000696 | 0.81 | 0.418 | -.00008 | .0001926 |
| IMPRCTfwup | -.0039241 | .0029853 | -1.31 | 0.189 | -.0097751 | .0019269 |
| IMPICER2 | .0000246 | .0000115 | 2.13 | 0.033 | 2.00e-06 | .0000472 |
| BNF7 | -.1610554 | .4803539 | -0.34 | 0.737 | -1.102532 | .7804211 |
| BNF13 | -.3224628 | .7241085 | -0.45 | 0.656 | -1.741689 | 1.096764 |
| IMPprevale~e | -.0000158 | 9.13e-06 | -1.73 | 0.083 | -.0000337 | 2.08e-06 |
| alternative | -.5311602 | .6109776 | -0.87 | 0.385 | -1.728654 | .6663339 |
| _cons | .9710208 | .6719289 | 1.45 | 0.148 | -.3459356 | 2.287977 |
| **not recomm~d** | | | | | | |
| IMPRCTsize | -.0002907 | .0001589 | -1.83 | 0.067 | -.0006022 | .0000208 |
| IMPRCTfwup | -.0095656 | .0036281 | -2.64 | 0.008 | -.0166765 | -.0024547 |
| IMPICER2 | .0000273 | .0000115 | 2.37 | 0.018 | 4.75e-06 | .0000498 |
| BNF7 | -1.674223 | .5820207 | -2.88 | 0.004 | -2.814962 | -.5334832 |
| BNF13 | -1.721107 | .918 | -1.87 | 0.061 | -3.520354 | .0781396 |
| IMPprevale~e | -.000015 | 8.75e-06 | -1.71 | 0.087 | -.0000321 | 2.15e-06 |
| alternative | -.9756176 | .6028344 | -1.62 | 0.106 | -2.157151 | .2059161 |
| _cons | 2.178807 | .6699368 | 3.25 | 0.001 | .8657553 | 3.491859 |

(decision==recommended is the base outcome)

**Table 3. SMC multinomial logistic regression results <u>with multiple imputation using ICE command</u>**

Multinomial logistic regression — Number of obs = 288; LR chi2(14) = 59.79; Prob > chi2 = 0.0000; Log likelihood = -269.35554; Pseudo R2 = 0.0999

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **restricted** | | | | | | |
| RCTsize | -.0000179 | .0000548 | -0.33 | 0.744 | -.0001252 | .0000895 |
| RCTfwup | -.0029613 | .0028239 | -1.05 | 0.294 | -.0084961 | .0025735 |
| ICER | 8.23e-06 | 3.21e-06 | 2.57 | 0.010 | 1.95e-06 | .0000145 |
| BNF7 | -.4062836 | .4904738 | -0.83 | 0.407 | -1.367595 | .5550274 |
| BNF13 | -.4085267 | .7258759 | -0.56 | 0.574 | -1.831217 | 1.014164 |
| prevalence | -3.97e-06 | 3.03e-06 | -1.31 | 0.190 | -9.90e-06 | 1.96e-06 |
| alternative | -.6032234 | .6058133 | -1.00 | 0.319 | -1.790596 | .5841489 |
| _cons | 1.37813 | .6212923 | 2.22 | 0.027 | .16042 | 2.595841 |
| **not recomm~d** | | | | | | |
| RCTsize | -.000246 | .0001048 | -2.35 | 0.019 | -.0004515 | -.0000405 |
| RCTfwup | -.0091805 | .0033508 | -2.74 | 0.006 | -.015748 | -.002613 |
| ICER | 9.70e-06 | 3.24e-06 | 2.99 | 0.003 | 3.35e-06 | .0000161 |
| BNF7 | -1.942678 | .5966816 | -3.26 | 0.001 | -3.112152 | -.7732035 |
| BNF13 | -1.894411 | .9185067 | -2.06 | 0.039 | -3.694651 | -.0941711 |
| prevalence | -4.91e-06 | 3.10e-06 | -1.58 | 0.114 | -.000011 | 1.17e-06 |
| alternative | -1.017091 | .5988571 | -1.70 | 0.089 | -2.190829 | .1566478 |
| _cons | 2.565645 | .6132337 | 4.18 | 0.000 | 1.363729 | 3.767561 |

(decision==recommended is the base outcome)

## Results

The results suggest that the various imputation techniques provided similar pseudo R-squared results across the models and similar pattern of size, direction and significance of effect. It was therefore felt to be appropriate, for the remainder of the multivariate analyses to impute missing values by using regression estimates of the missing value.

## References

UCLA . Multiple Imputation Using ICE. UCLA: Academic Technology Services, Statistical Consulting Group. from http://www.ats.ucla.edu/stat/stata/library/ice.htm (accessed January 5th 2011).

# G. Implications of assuming ordinality of the outcome variable on multivariate analyses

## Overview

In performing the multivariate analysis of coverage decisions issued by NICE, SMC, CFH and HAS, an important consideration from a methodological standpoint is whether ordinality should be assumed, as this has implications for the statistical methods that should be applied. A categorical variable can be considered ordinal if there is a 'natural' ordering in the outcome that can be identified. Dakin et al (2006), in their econometric analysis of NICE decision-making, used a three-category approach, and argued that an assumption of ordinality was not appropriate as there was no consensus on the 'direction' of the ordering. In other words, depending on whether the aim of the appraisal decision was to define volume (i.e. number of patients that can access the therapy), total drug cost for the technology (budget impact) or maximization of patient/healthcare system outcome can lead to different ordering and consideration of appraisal decisions. For example, if the main objective of the coverage decision was to control cost, then it could be expected that those decisions which restrict or not recommend access may be considered as preferable to a recommendation. The opposite is true if we consider patient access to medications as the main objective of coverage decisions: in this case, a coverage decision that covers 100% of the eligible population, less than 100% or 0% could be considered. In the base case analysis, multinomial logit regression was assumed to be appropriate. Here, multivariate models were run for each of the four HTA bodies in which ordinality of the outcome variable was assumed. The results of these analyses are compared with the base-case analyses and are presented below.

## Ordinal Logistic Regression method

A well known ordinal logistic regression model is the proportional odds model. It assumes that the relationship between each pair of outcome groups is the same. This is a significant assumption and is highly restrictive – in many cases the coefficient for a variable in one group may be different from that in the lower category group (Williams 2006). The proportional odds model can only be applied if the assumption of proportional odds is not violated. To assess whether this is the case, two tests were performed: A likelihood ratio test (using the omodel command by Wolfe and Gould's

(1998)) and a Brant test (Long and Freese) were performed to assess whether the proportional odds model applied was appropriate or not – i.e. whether the assumption of proportional odds was violated. The Brant test provides both a global test of whether any variable violates the parallel-lines assumption, as well as tests of the assumption for each variable separately (Williams 2006).

In the event that the assumption of proportional odds is violated, several options exist that provide alternative modelling options. The first would be to revert to a non-ordinal model, such as the multinomial logistic regression model. This is the model that was used in the base case analyses of this thesis, and therefore will not be repeated in this sensitivity analysis. Another option to the proportional odds model would be to use a generalized ordered logistic model. Therefore, a generalized ordered logistic model was run. To run such a model, the gologit2 command developed by Williams (2006) was used. The advantage of gologit2 is that, rather than assuming that the proportional odds assumption is violated for all variables, it assesses those variables for which this is actually the case, and fits partial proportional odds models to those variables where the assumption holds, and generalized linear models where it does not hold – in other words, the restriction of the proportional odds assumption is relaxed for those variables that violate the assumption only (Williams 2006).

When interpreting the results of the generalized ordered logit model it is important to bear in mind that, while the output resembles that obtained from multinomial logistic regression, the reference case is not the same. In the generalized ordered logit model, the reference case is a group of outcome variables. The first panel of output compares recommended technologies (category 1) with both restricted (cat. 2) and not-recommended technologies (cat. 3), while the second panel compares restricted (cat. 2) and recommended (cat. 1) categories with not recommended technologies (cat. 3, reference) (Williams 2006). Within the model output, positive coefficients signal that an increase in the unit of the explanatory variable increased the odds that the technology will be in a higher outcome category (category 3 is the highest category representing non recommendation). The opposite is true for negative coefficients.

**Results**

*NICE*

When the proportional odds ordinal logistic regression model was fitted to the data, both the likelihood ratio test and the Brant test were significant (at the 0.05) level, meaning that the assumption of proportional odds between the outcome categories was violated. Therefore, a generalized ordered logistic model was run. Table 1 provides the base-case model for NICE coverage decisions.

**Table 1. Multivariate analysis of NICE coverage decisions 2004-2009: base case model results (n=118)**

| Restricted | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -1.567512 | 0.006 | -2.678506 | -0.456519 |
| ICER | 0.0000484 | 0.009 | 0.0000123 | 0.0000846 |
| Number of technologies appraised simultaneously | 0.4891155 | 0.005 | 1.44E-01 | 0.8338142 |
| Year of Appraisal | 0.3579406 | 0.072 | -0.0317402 | 0.7476213 |
| Constant | -718.9729 | 0.071 | -1500.761 | 62.81519 |
| Not Recommended | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -2.009114 | 0.016 | -3.636783 | -0.3814446 |
| ICER | 0.0000865 | 0.000 | 0.0000419 | 0.0001311 |
| Number of technologies appraised simultaneously | 0.1199482 | 0.645 | -3.91E-01 | 0.630705 |
| Year of Appraisal | 0.6721141 | 0.028 | 0.0733939 | 1.270834 |
| Constant | -1351.138 | 0.028 | -2552.896 | -149.3792 |

Note: Recommended technologies are the reference case

**Table 2. NICE dataset - Generalized Ordered Logit Model (proportional odds constraint relaxed for all variables)**

```
Generalized Ordered Logit Estimates              Number of obs   =        118
                                                 LR chi2(8)      =      64.30
                                                 Prob > chi2     =     0.0000
Log likelihood = -79.569463                      Pseudo R2       =     0.2878
```

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **Recommended** | | | | | | |
| ICER | .0000663 | .0000191 | 3.47 | 0.001 | .0000288 | .0001038 |
| superior_yes | -1.660658 | .5602506 | -2.96 | 0.003 | -2.758729 | -.5625873 |
| no_drugs | .5187612 | .181614 | 2.86 | 0.004 | .1628043 | .874718 |
| year | .4257265 | .2023483 | 2.10 | 0.035 | .0291311 | .8223218 |
| _cons | -855.2074 | 406.0203 | -2.11 | 0.035 | -1650.993 | -59.42216 |
| **Restricted** | | | | | | |
| ICER | .000071 | .0000217 | 3.28 | 0.001 | .0000286 | .0001134 |
| superior_yes | -1.136288 | .727402 | -1.56 | 0.118 | -2.56197 | .2893934 |
| no_drugs | -.2336803 | .2087955 | -1.12 | 0.263 | -.642912 | .1755513 |
| year | .2726342 | .2535441 | 1.08 | 0.282 | -.224303 | .7695715 |
| _cons | -550.1457 | 508.9175 | -1.08 | 0.280 | -1547.606 | 447.3143 |

Generalized ordinal logit models were implemented – one model (Table 2) the proportional odds constraint relaxed was for all variables, while in the other model the

proportional odds assumption relaxed only for variables that violated the assumption (Table 3). The results of the model are shown in Table 2 – and provide very similar results to the base case analyses run using a multinomial logistic regression model. In the first panel, recommended technologies are compared with restricted and not recommended technologies. An increase in the ICER increases the odds of restriction and non-recommendation, as does an increase in the year of appraisal and the number of technologies appraised simultaneously. The demonstration of superiority within the RCT decreases the odds of restriction or non-recommendation. These results are in line with those observed in the base case analysis. In the second panel, recommended and restricted technologies were compared with non-recommended technologies (reference). In this panel, an increase in the ICER increased the odds of non-recommendation, as did the year of appraisal. The demonstration of clinical superiority and an increase in the number of technologies appraised simultaneously decreased the odds of non-recommendation, increasing the odds of either recommendation or restriction.

**Table 3. NICE Generalized Ordered Logit Regression (proportional odds assumption relaxed only for variables that violate the assumption)**

Generalized Ordered Logit Estimates — Number of obs = 118; Wald chi2(5) = 34.37; Prob > chi2 = 0.0000; Log likelihood = -79.867881; Pseudo R2 = 0.2851

( 1) [Recommended]ICER - [Restricted]ICER = 0
( 2) [Recommended]year - [Restricted]year = 0
( 3) [Recommended]superior_yes - [Restricted]superior_yes = 0

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **Recommended** | | | | | | |
| ICER | .0000691 | .0000152 | 4.54 | 0.000 | .0000393 | .000099 |
| superior_yes | -1.47203 | .4768137 | -3.09 | 0.002 | -2.406568 | -.5374923 |
| no_drugs | .5114306 | .1765594 | 2.90 | 0.004 | .1653805 | .8574806 |
| year | .363303 | .1648912 | 2.20 | 0.028 | .0401222 | .6864838 |
| _cons | -730.0836 | 330.8916 | -2.21 | 0.027 | -1378.619 | -81.54787 |
| **Restricted** | | | | | | |
| ICER | .0000691 | .0000152 | 4.54 | 0.000 | .0000393 | .000099 |
| superior_yes | -1.47203 | .4768137 | -3.09 | 0.002 | -2.406568 | -.5374923 |
| no_drugs | -.2428722 | .2036811 | -1.19 | 0.233 | -.6420799 | .1563355 |
| year | .363303 | .1648912 | 2.20 | 0.028 | .0401222 | .6864838 |
| _cons | -731.9488 | 330.9301 | -2.21 | 0.027 | -1380.56 | -83.33773 |

*SMC*

When the proportional odds ordinal logistic regression model was fitted to the SMC data, both the likelihood ratio test and the Brant test were significant (at the 0.05) level, meaning that the assumption of proportional odds between the outcome categories was violated. Therefore, a generalized ordered logistic model was run.

Table 4 SMC Base case multinomial logistic regression

| Restricted Technologies | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| RCT size | 0.0000563 | 0.418 | -0.00008 | 0.0001926 |
| RCT duration of follow-up | -0.0039241 | 0.189 | -0.0097751 | 0.0019269 |
| ICER | 0.0000246 | 0.033 | 0.000002 | 0.0000472 |
| Infectious Diseases | -0.1610554 | 0.737 | -1.102532 | 0.7804211 |
| Skin Diseases | -0.3224628 | 0.656 | -1.741689 | 1.096764 |
| Disease prevalence | -0.0000158 | 0.083 | -0.0000337 | 2.08E-06 |
| Presence of alternative therapy | -0.5311602 | 0.385 | -1.728654 | 0.6663339 |
| Constant | 0.9710208 | 0.148 | -0.3459356 | 2.287977 |
| Restricted Technologies | Log Odds | P value | 95% Conf. Interval | |
| RCT size | -0.0002907 | 0.067 | -0.0006022 | 0.0000208 |
| RCT duration of follow-up | -0.0095656 | 0.008 | -0.0166765 | -0.0024547 |
| ICER | 0.0000273 | 0.018 | 4.75E-06 | 0.0000498 |
| Infectious Diseases | -1.674223 | 0.004 | -2.814962 | -0.5334832 |
| Skin Diseases | -1.721107 | 0.061 | -3.520354 | 0.0781396 |
| Disease prevalence | -0.000015 | 0.087 | -0.0000321 | 2.15E-06 |
| Presence of alternative therapy | -0.9756176 | 0.106 | -2.157151 | 0.2059161 |
| Constant | 2.178807 | 0.001 | 0.8657553 | 3.491859 |

The resulting model for this sensitivity analysis is shown below. The pseudo R squared for this model is similar to that of the base case model using multinomial logistic regression techniques (0.0981 vs. 0.1063). The results of the model are shown in Table 5 – and provide generally similar results to the base case analyses run using a multinomial logistic regression model (Table 4). In the first panel, recommended technologies are compared with restricted and not recommended technologies. An increase in RCT size increases the odds of a recommendation, as does an increase in the duration of follow-up. An increase in the ICER increases the odds of restriction or non recommendation, although in this sensitivity analysis the statistical significance of the effect of the ICER is borderline (p=0.102). Technologies indicated for the treatment of infectious diseases or skin diseases increase the odds of recommendation, as was also

seen in the base-case analyses. An increasing prevalence increased the odds of recommendation. Its effect in the lower panel of results was not significant. Table 6 provides results for the generalized ordinal logit model in which proportional odds assumption relaxed only for variables that violate the assumption. The resulting pseudo R-squared is 0.0811, lower than that observed in the base case model (0.1063), although the direction and size of effect observed in this model is similar to that observed in the previous generalized model. .

Table 5 SMC Generalized Ordered Logit Regression (proportional odds assumption relaxed for all variables)

Generalized Ordered Logit Estimates — Number of obs = 288; LR chi2(14) = 58.73; Prob > chi2 = 0.0000; Log likelihood = -269.88325; Pseudo R2 = 0.0981

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **recommended** | | | | | | |
| IMPRCTsize | -.0000316 | .0000616 | -0.51 | 0.608 | -.0001524 | .0000892 |
| IMPRCTfwup | -.0062476 | .0026463 | -2.36 | 0.018 | -.0114342 | -.0010609 |
| IMPICER2 | .0000163 | 9.96e-06 | 1.63 | 0.102 | -3.25e-06 | .0000358 |
| BNF7 | -.8027816 | .4520825 | -1.78 | 0.076 | -1.688847 | .0832838 |
| BNF13 | -1.003322 | .6718258 | -1.49 | 0.135 | -2.320076 | .3134322 |
| IMPprevale~e | -.0000136 | 6.95e-06 | -1.96 | 0.050 | -.0000272 | 1.32e-08 |
| alternative | -.8496164 | .5720232 | -1.49 | 0.137 | -1.970761 | .2715285 |
| _cons | 2.509094 | .6185565 | 4.06 | 0.000 | 1.296745 | 3.721442 |
| **restricted** | | | | | | |
| IMPRCTsize | -.0002936 | .0001384 | -2.12 | 0.034 | -.0005648 | -.0000224 |
| IMPRCTfwup | -.0071538 | .0028702 | -2.49 | 0.013 | -.0127794 | -.0015283 |
| IMPICER2 | 4.95e-06 | 4.41e-06 | 1.12 | 0.261 | -3.68e-06 | .0000136 |
| BNF7 | -1.661123 | .4849069 | -3.43 | 0.001 | -2.611523 | -.7107234 |
| BNF13 | -1.391018 | .8042916 | -1.73 | 0.084 | -2.967401 | .1853643 |
| IMPprevale~e | -8.39e-06 | 6.73e-06 | -1.25 | 0.213 | -.0000216 | 4.81e-06 |
| alternative | -.5606204 | .3587356 | -1.56 | 0.118 | -1.263729 | .1424883 |
| _cons | .9601263 | .3889691 | 2.47 | 0.014 | .197761 | 1.722492 |

Table 6. SMC Generalized Ordered Logit Regression (proportional odds assumption relaxed only for variables that violate the assumption)

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **recommended** | | | | | | |
| IMPRCTsize | -.0000961 | .0000467 | -2.06 | 0.040 | -.0001877 | -4.59e-06 |
| IMPRCTfwup | -.0072192 | .0022261 | -3.24 | 0.001 | -.0115823 | -.0028561 |
| IMPICER2 | 6.98e-06 | 4.91e-06 | 1.42 | 0.155 | -2.64e-06 | .0000166 |
| BNF7 | -1.241631 | .3584794 | -3.46 | 0.001 | -1.944238 | -.5390246 |
| BNF13 | -1.24658 | .5908885 | -2.11 | 0.035 | -2.4047 | -.0884596 |
| IMPprevale~e | -.0000103 | 6.03e-06 | -1.71 | 0.087 | -.0000222 | 1.49e-06 |
| alternative | -.5743616 | .3324418 | -1.73 | 0.084 | -1.225935 | .0772123 |
| _cons | 2.642918 | .3944229 | 6.70 | 0.000 | 1.869864 | 3.415973 |
| **restricted** | | | | | | |
| IMPRCTsize | -.0000961 | .0000467 | -2.06 | 0.040 | -.0001877 | -4.59e-06 |
| IMPRCTfwup | -.0072192 | .0022261 | -3.24 | 0.001 | -.0115823 | -.0028561 |
| IMPICER2 | 6.98e-06 | 4.91e-06 | 1.42 | 0.155 | -2.64e-06 | .0000166 |
| BNF7 | -1.241631 | .3584794 | -3.46 | 0.001 | -1.944238 | -.5390246 |
| BNF13 | -1.24658 | .5908885 | -2.11 | 0.035 | -2.4047 | -.0884596 |
| IMPprevale~e | -.0000103 | 6.03e-06 | -1.71 | 0.087 | -.0000222 | 1.49e-06 |
| alternative | -.5743616 | .3324418 | -1.73 | 0.084 | -1.225935 | .0772123 |
| _cons | .7786995 | .3621462 | 2.15 | 0.032 | .068906 | 1.488493 |

*CFH*

When the proportional odds ordinal logistic regression model was fitted to the data, both the likelihood ratio test and the Brant test were significant (at the 0.05) level, meaning that the assumption of proportional odds between the outcome categories was violated. Therefore, a generalized ordered logistic model was run.

**Table 7. Multivariate analysis of CFH coverage decisions 2004-2009: base case model results**

| **Restricted** | **Log Odds** | **P value** | **95% Confidence Interval** | |
|---|---|---|---|---|
| Use of active comparator in RCT | -0.4926263 | 0.239 | -1.312877 | 0.3276248 |
| Demonstrated clinical superiority in RCT | -0.8223267 | 0.022 | -1.52789 | -0.1167636 |
| Budgetary Impact | 0.0068947 | 0.051 | -0.0000311 | 0.0138206 |
| Cancer therapy | -1.593124 | 0.000 | -2.467556 | -0.7186923 |
| Therapies for cardiovascular diseases | 1.401083 | 0.017 | 0.2532709 | 2.548895 |
| Therapies for obstetrics, gynaecology, and urinary-tract disorders | 2.396135 | 0.032 | 0.2027481 | 4.589521 |
| Prevalence of target population | -0.00000135 | 0.091 | -0.0000029 | 2.15E-07 |
| Patient Submission | -0.149162 | 0.879 | -2.075549 | 1.777225 |
| Lack of data on duration of RCT | 0.2441296 | 0.560 | -0.5760478 | 1.064307 |
| Constant | 0.2812028 | 0.459 | -0.4628566 | 1.025262 |
| **Not Recommended** | **Log Odds** | **P value** | **95% Confidence Interval** | |
| Use of active comparator in RCT | -2.541106 | 0.000 | -3.839421 | -1.242791 |
| Demonstrated clinical superiority in RCT | -1.853915 | 0.000 | -2.818847 | -0.8889837 |
| Budgetary Impact | 0.0047054 | 0.217 | -0.0027627 | 0.0121735 |
| Cancer therapy | -0.6968874 | 0.177 | -1.70854 | 0.3147655 |
| Therapies for cardiovascular diseases | 0.823405 | 0.338 | -0.8608332 | 2.507643 |
| Therapies for obstetrics, gynaecology, and urinary-tract disorders | 2.197455 | 0.141 | -0.7261041 | 5.121014 |
| Prevalence of target population | -1.64E-08 | 0.982 | -1.42E-06 | 1.38E-06 |
| Patient Submission | 1.786073 | 0.032 | 0.1576834 | 3.414462 |
| Lack of data on duration of RCT | 1.775784 | 0.001 | 0.7376478 | 2.813921 |
| Constant | 0.0663455 | 0.878 | -0.7782347 | 0.9109256 |

The results of the model are shown in Table 8 – and provide generally similar results to the base case analyses run using a multinomial logistic regression model (Table 7). The model results for this sensitivity analysis yield a slightly higher pseudo R-squared than in the base case model (0.1929 vs. 0.1725). In the first panel, recommended technologies are compared with restricted and not recommended technologies. An increase in the proportion of trials with an active comparator, the demonstration of superiority, and cancer therapies increase the odds of recommendation, as was observed in the base-case analysis. Technologies indicated for the treatment of cardiovascular disorders or for obstetrics/gynaecology disorders, as well as the presence of a patient submission and the lack of information on the trial duration increased the odds of

restriction or non-recommendation, also consistent with the base case analysis. Table 8 provides results for the generalized ordinal logit model in which proportional odds assumption relaxed only for variables that violate the assumption. The resulting pseudo R-squared is 0.1646, similar to that observed in the base case model (0.1725), and the direction and size of effect observed in this model is similar to that observed in the previous generalized model.

**Table 8. CFH Generalized Ordered Logit Regression (proportional odds assumption relaxed for all variables)**

Generalized Ordered Logit Estimates — Number of obs = 256; LR chi2(18) = 98.82; Prob > chi2 = 0.0000; Log likelihood = -206.74479; Pseudo R2 = 0.1929

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **Recommended** | | | | | | |
| IMPActive_~p | -.8984875 | .4089198 | -2.20 | 0.028 | -1.699956 | -.0970195 |
| IMPsuperio~s | -1.000851 | .3566156 | -2.81 | 0.005 | -1.699805 | -.3018978 |
| IMPBIM | .0064387 | .0036572 | 1.76 | 0.078 | -.0007292 | .0136066 |
| CANCER | -1.21891 | .3625657 | -3.36 | 0.001 | -1.929526 | -.5082944 |
| BNF1 | 1.421132 | .5903941 | 2.41 | 0.016 | .2639811 | 2.578284 |
| BNF11 | 2.417799 | 1.112583 | 2.17 | 0.030 | .2371766 | 4.598422 |
| RCTfwup_M | .6866893 | .3764034 | 1.82 | 0.068 | -.0510478 | 1.424426 |
| IMPprevale~e | -8.71e-07 | 5.84e-07 | -1.49 | 0.136 | -2.02e-06 | 2.74e-07 |
| patient_sub | .7976289 | .7711926 | 1.03 | 0.301 | -.7138808 | 2.309139 |
| _cons | .7044065 | .3685162 | 1.91 | 0.056 | -.017872 | 1.426685 |
| **Restricted** | | | | | | |
| IMPActive_~p | -3.081604 | .755256 | -4.08 | 0.000 | -4.561878 | -1.601329 |
| IMPsuperio~s | -1.710759 | .4928227 | -3.47 | 0.001 | -2.676674 | -.7448441 |
| IMPBIM | -.0039186 | .0040164 | -0.98 | 0.329 | -.0117905 | .0039534 |
| CANCER | -.1113597 | .5143169 | -0.22 | 0.829 | -1.119402 | .8966828 |
| BNF1 | .0195211 | .7024361 | 0.03 | 0.978 | -1.357228 | 1.39627 |
| BNF11 | .9296203 | 1.163985 | 0.80 | 0.424 | -1.351749 | 3.21099 |
| RCTfwup_M | 2.071838 | .5372991 | 3.86 | 0.000 | 1.018752 | 3.124925 |
| IMPprevale~e | 2.34e-06 | 1.02e-06 | 2.30 | 0.021 | 3.46e-07 | 4.34e-06 |
| patient_sub | 2.475603 | .803122 | 3.08 | 0.002 | .9015124 | 4.049693 |
| _cons | -.6668427 | .3801655 | -1.75 | 0.079 | -1.411953 | .078268 |

Table **9. Generalized Ordered Logit Model(proporitional odds constraint relaxed only for variables where assumption is violated, using 'autofit' command)**

| decision | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **Recommended** | | | | | | |
| IMPActive_~p | -1.219262 | .3878347 | -3.14 | 0.002 | -1.979404 | -.4591198 |
| IMPsuperio~s | -1.268082 | .3186665 | -3.98 | 0.000 | -1.892657 | -.6435067 |
| IMPBIM | .0012359 | .001163 | 1.06 | 0.288 | -.0010435 | .0035153 |
| CANCER | -1.251005 | .3585961 | -3.49 | 0.000 | -1.95384 | -.5481695 |
| BNF1 | .8584665 | .4500128 | 1.91 | 0.056 | -.0235423 | 1.740475 |
| BNF11 | 1.731503 | .7360465 | 2.35 | 0.019 | .2888784 | 3.174128 |
| RCTfwup_M | .7963189 | .3672564 | 2.17 | 0.030 | .0765095 | 1.516128 |
| IMPprevale~e | -6.64e-07 | 5.59e-07 | -1.19 | 0.235 | -1.76e-06 | 4.32e-07 |
| patient_sub | .4697717 | .7652413 | 0.61 | 0.539 | -1.030074 | 1.969617 |
| _cons | 1.146864 | .3340843 | 3.43 | 0.001 | .4920711 | 1.801657 |
| **Restricted** | | | | | | |
| IMPActive_~p | -2.839025 | .7090808 | -4.00 | 0.000 | -4.228798 | -1.449252 |
| IMPsuperio~s | -1.268082 | .3186665 | -3.98 | 0.000 | -1.892657 | -.6435067 |
| IMPBIM | .0012359 | .001163 | 1.06 | 0.288 | -.0010435 | .0035153 |
| CANCER | -.1595909 | .495196 | -0.32 | 0.747 | -1.130157 | .8109754 |
| BNF1 | .8584665 | .4500128 | 1.91 | 0.056 | -.0235423 | 1.740475 |
| BNF11 | 1.731503 | .7360465 | 2.35 | 0.019 | .2888784 | 3.174128 |
| RCTfwup_M | 1.985576 | .4832085 | 4.11 | 0.000 | 1.038505 | 2.932648 |
| IMPprevale~e | 1.46e-06 | 9.73e-07 | 1.50 | 0.134 | -4.51e-07 | 3.36e-06 |
| patient_sub | 2.641552 | .8020039 | 3.29 | 0.001 | 1.069653 | 4.21345 |
| _cons | -1.13452 | .3773007 | -3.01 | 0.003 | -1.874016 | -.3950244 |

*HAS*

When the proportional odds ordinal logistic regression model was fitted to the HAS data, the likelihood ratio test was found to be statistically significant at the 0.05 level, but the Brant test was not found to be statistically significant (p= 0.147). Thus the diagnostic tests provided mixed feedback as to whether the ordinal logit regression model based on proportional odds was appropriate or not for the data set. To this end, both the ordinal logit as well as the generalized ordered logistic model are presented.

**Table 10. HAS Base Case analysis – multinomial logistic regression**

| ASMR III-IV | Log Odds | P value | 95% Conf. Interval | |
|---|---|---|---|---|
| Clinical superiority demonstrated in RCT | -0.4204505 | 0.277 | -1.178813 | 0.337912 |
| Disease prevalence | 0.000000489 | 0.113 | -1.15E-07 | 0.00000109 |
| Orphan designation status | -1.271573 | 0.010 | -2.24461 | -0.2985352 |
| Central nervous system | 2.870115 | 0.007 | 0.786877 | 4.953353 |
| Infections | 2.311445 | 0.036 | 0.1485392 | 4.47435 |
| Musculoskeletal and joint diseases | -1.58E+00 | 0.002 | -2.61E+00 | -5.57E-01 |
| Healthcare expenditure on pharmaceuticals | 9.12E-03 | 0.273 | -7.20E-03 | 2.54E-02 |
| Missing data on prevalence of disease | -2.14573 | 0.001 | -3.376858 | -0.914602 |
| Constant | -3.449974 | 0.419 | -11.81306 | 4.913112 |
| ASMR V | Log Odds | P value | 95% Conf. Interval | |
| Clinical superiority demonstrated in RCT | -1.349468 | 0.001 | -2.167562 | -0.5313733 |

| | | | | |
|---|---|---|---|---|
| Disease prevalence | 0.000000875 | 0.005 | 0.000000265 | 0.00000149 |
| Orphan designation status | -3.522852 | 0.000 | -5.18217 | -1.863535 |
| Central nervous system | 3.170685 | 0.003 | 1.070034 | 5.271336 |
| Infections | 2.213558 | 0.046 | 0.0441185 | 4.382997 |
| Musculoskeletal and joint diseases | -2.58E+00 | 0.000 | -3.87E+00 | -1.28E+00 |
| Healthcare expenditure on pharmaceuticals | 2.39E-02 | 0.008 | 6.33E-03 | 4.15E-02 |
| Missing data on prevalence of disease | -2.746339 | 0.000 | -3.980416 | -1.512263 |
| Constant | -10.67378 | 0.021 | -19.72712 | -1.620434 |

The results of the model are shown in Table 10. Having modelled HAS decision-making using the five ASMR categories means that a straightforward comparison with the base case analysis is challenging. The model results for this sensitivity analysis yield a slightly higher pseudo R-squared than in the base case model (0.1862 vs. 0.1878). Thus assuming the ordinality assumption and modelling using a five-category outcome variable does not increase the ability for the model to explain a larger percentage of HAS decision-making, suggesting that a multinomial approach using a 3-category variable may be appropriate. The results however, of this sensitivity analysis, reveal more detail in how the factors behave within the different comparisons, as highlighted by the four panels displayed in Table 11. In the first panel, ASMR I technologies are compared with ASMR II-V technologies. In this particular panel two variables demonstrate a statistically significant effect: orphan designation increases the odds of an ASMR I. Oddly, those technologies for which no information was available on the prevalence of the disease also increased the odds of an ASMR I relative to ASMR II-V. In the second panel, technologies with ASMR I-II are compared with ASMR III-V technologies, similar to the sensitivity analysis in which binary category was used, although in that case, ordinality was not assumed. In this particular panel, all variables have statistically significant effects with the exception of trial duration and technologies indicated for infectious diseases, which do not demonstrate statistical significance. Consistent with results in the base case analysis, demonstration of superiority, a longer trial duration, orphan designation, and indication for the treatment of musculoskeletal diseases increases the odds of an ASMR I-II relative to technologies with an ASMR III-V. The remaining panels (3 and 4) reflect closely the results obtained in the base case model, while providing a more detailed view on the behaviour of each explanatory variable within the various ASMR ratings. The results obtained from the ordinal logit model, which assumes proportional odds, are very similar to those

obtained in the base case multinomial logistic regression model, although the pseudo R-squared is lower than the base case model (0.1279 vs. 0.1878).

Table 11 HAS Generalized Ordered Logit Regression (proportional odds assumption relaxed for all variables), using 5 ASMR categories

Generalized Ordered Logit Estimates — Number of obs = 315; LR chi2(36) = 162.81; Prob > chi2 = 0.0000; Log likelihood = -355.76654; Pseudo R2 = 0.1862

| asmr | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| **1** | | | | | | |
| IMPsuperio~s | -.8675846 | .8529147 | -1.02 | 0.309 | -2.539267 | .8040976 |
| IMP_fwup_RCT | .004596 | .0084366 | 0.54 | 0.586 | -.0119393 | .0211313 |
| IMP_preval~e | 5.04e-07 | 5.16e-07 | 0.98 | 0.329 | -5.07e-07 | 1.51e-06 |
| orphan | -1.77837 | .7525145 | -2.36 | 0.018 | -3.253271 | -.3034689 |
| BNF2 | 17.4004 | 2530.209 | 0.01 | 0.995 | -4941.717 | 4976.518 |
| BNF7 | 15.6971 | 2537.926 | 0.01 | 0.995 | -4958.547 | 4989.942 |
| BNF9 | 12.69895 | 550.5427 | 0.02 | 0.982 | -1066.345 | 1091.743 |
| drug_exp | .0044086 | .0175877 | 0.25 | 0.802 | -.0300627 | .0388799 |
| prevalence_M | -2.313847 | 1.080178 | -2.14 | 0.032 | -4.430958 | -.1967366 |
| _cons | 1.346192 | 8.845789 | 0.15 | 0.879 | -15.99124 | 18.68362 |
| **2** | | | | | | |
| IMPsuperio~s | -.7725188 | .389363 | -1.98 | 0.047 | -1.535656 | -.0093813 |
| IMP_fwup_RCT | -.0037046 | .0028905 | -1.28 | 0.200 | -.00937 | .0019607 |
| IMP_preval~e | 5.90e-07 | 2.81e-07 | 2.10 | 0.036 | 3.98e-08 | 1.14e-06 |
| orphan | -1.864821 | .4813334 | -3.87 | 0.000 | -2.808218 | -.9214254 |
| BNF2 | 2.784682 | 1.036082 | 2.69 | 0.007 | .7539978 | 4.815365 |
| BNF7 | 1.646184 | 1.057101 | 1.56 | 0.119 | -.4256956 | 3.718063 |
| BNF9 | -1.937744 | .4870652 | -3.98 | 0.000 | -2.892374 | -.9831137 |
| drug_exp | .0139212 | .0077513 | 1.80 | 0.072 | -.001271 | .0291135 |
| prevalence_M | -2.122849 | .5699047 | -3.72 | 0.000 | -3.239842 | -1.005857 |
| _cons | -4.794615 | 4.025102 | -1.19 | 0.234 | -12.68367 | 3.094439 |
| **3** | | | | | | |
| IMPsuperio~s | -.826269 | .289044 | -2.86 | 0.004 | -1.392785 | -.2597532 |
| IMP_fwup_RCT | -.0057788 | .0028616 | -2.02 | 0.043 | -.0113874 | -.0001701 |
| IMP_preval~e | 4.63e-07 | 2.00e-07 | 2.31 | 0.021 | 7.08e-08 | 8.55e-07 |
| orphan | -1.806465 | .5011118 | -3.60 | 0.000 | -2.788626 | -.8243043 |
| BNF2 | 2.214457 | .5754387 | 3.85 | 0.000 | 1.086618 | 3.342296 |
| BNF7 | .4197093 | .4375743 | 0.96 | 0.337 | -.4379205 | 1.277339 |
| BNF9 | -1.656029 | .5305292 | -3.12 | 0.002 | -2.695847 | -.6162109 |
| drug_exp | .0252388 | .0060627 | 4.16 | 0.000 | .0133561 | .0371214 |
| prevalence_M | -1.478545 | .4197982 | -3.52 | 0.000 | -2.301334 | -.6557555 |
| _cons | -11.87794 | 3.10484 | -3.83 | 0.000 | -17.96332 | -5.792568 |
| **4** | | | | | | |
| IMPsuperio~s | -.9607761 | .2807114 | -3.42 | 0.001 | -1.51096 | -.4105918 |
| IMP_fwup_RCT | -.0021929 | .0026662 | -0.82 | 0.411 | -.0074185 | .0030327 |
| IMP_preval~e | 5.33e-07 | 1.79e-07 | 2.98 | 0.003 | 1.83e-07 | 8.84e-07 |
| orphan | -2.422181 | .769751 | -3.15 | 0.002 | -3.930866 | -.9134971 |
| BNF2 | .6311121 | .3472323 | 1.82 | 0.069 | -.0494507 | 1.311675 |
| BNF7 | .2899447 | .3934601 | 0.74 | 0.461 | -.4812229 | 1.061112 |
| BNF9 | -1.038595 | .591875 | -1.75 | 0.079 | -2.198649 | .1214584 |
| drug_exp | .0153856 | .0058062 | 2.65 | 0.008 | .0040057 | .0267656 |
| prevalence_M | -1.012878 | .3944723 | -2.57 | 0.010 | -1.786029 | -.2397263 |
| _cons | -7.787468 | 3.003142 | -2.59 | 0.010 | -13.67352 | -1.901418 |

Table 12 Ordinal Logistic Regression, assuming proportional odds

Ordered logit estimates — Number of obs = 315; LR chi2(9) = 111.84; Prob > chi2 = 0.0000; Log likelihood = -381.24898; Pseudo R2 = 0.1279

| asmr | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| IMPsuperio~s | -.8657282 | .2335222 | -3.71 | 0.000 | -1.323423 | -.4080332 |
| IMP_fwup_RCT | -.0037254 | .001996 | -1.87 | 0.062 | -.0076374 | .0001867 |
| IMP_preval~e | 5.11e-07 | 1.58e-07 | 3.24 | 0.001 | 2.02e-07 | 8.21e-07 |
| orphan | -1.864357 | .3675365 | -5.07 | 0.000 | -2.584716 | -1.143999 |
| BNF2 | 1.033565 | .3133411 | 3.30 | 0.001 | .4194279 | 1.647702 |
| BNF7 | .5002555 | .3625853 | 1.38 | 0.168 | -.2103987 | 1.21091 |
| BNF9 | -1.467647 | .3850269 | -3.81 | 0.000 | -2.222286 | -.7130081 |
| drug_exp | .0175786 | .0049441 | 3.56 | 0.000 | .0078882 | .0272689 |
| prevalence_M | -1.467968 | .3750687 | -3.91 | 0.000 | -2.203089 | -.7328468 |
| _cut1 | 4.581367 | 2.547416 | (Ancillary parameters) | | | |
| _cut2 | 6.633968 | 2.542549 | | | | |
| _cut3 | 7.958039 | 2.556279 | | | | |
| _cut4 | 8.979997 | 2.568308 | | | | |

## References

Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal* 6(1): 58–82.

# H. Chapter 8 appendices - Pooled analyses, descriptive statistics

**Table 1. Descriptive Statistics of pooled sample of coverage decisions from NICE, SMC, CFH and HAS (n= 977) by coverage decision**

| | Recommended | | | Restricted | | | Not Recommended | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | | |
| NICE | 12% | 8% | 16% | 18% | 14% | 22% | 5% | 3% | 8% | <0.01 | 2 |
| SMC | 20% | 15% | 25% | 27% | 22% | 31% | 40% | 35% | 45% | <0.01 | 2 |
| CFH | 49% | 43% | 55% | 23% | 18% | 27% | 12% | 9% | 16% | <0.01 | 2 |
| HAS | 19% | 14% | 24% | 32% | 27% | 37% | 43% | 38% | 48% | <0.01 | 2 |
| Number of RCTs considered in decision | 2.7 | 2.1 | 3.3 | 3.4 | 2.9 | 3.8 | 2.6 | 2.3 | 2.9 | <0.05 | 3 |
| Size of population included in RCTs | 967 | 687 | 1247 | 1251 | 914 | 1587 | 830 | 607 | 1053 | NS | 1 |
| Statistically Significant results - yes | 49% | 43% | 55% | 42% | 37% | 47% | 39% | 33% | 44% | <0.05 | 2 |
| no | 15% | 11% | 20% | 16% | 13% | 20% | 18% | 14% | 23% | <0.01 | 2 |
| inconsistent | 20% | 15% | 25% | 32% | 27% | 37% | 29% | 24% | 34% | NS | 2 |
| Length/extent of follow-up in RCT | 58.4 | 50.4 | 66.3 | 49.1 | 42.9 | 55.4 | 41.4 | 34.6 | 48.2 | <0.01 | 1 |
| Use of active comparator | 52% | 46% | 58% | 45% | 40% | 49% | 43% | 38% | 48% | NS | 1 |
| Number of observational studies considered in guidance | 0.6 | 0.3 | 0.8 | 0.3 | 0.2 | 0.4 | 1.2 | -0.3 | 2.7 | <0.05 | 3 |
| Consideration of Cost Utility Analysis in guidance | 31% | 26% | 37% | 39% | 34% | 44% | 37% | 31% | 42% | NS | 2 |
| Incremental Cost-effectiveness ratio of | £20,151 | £12,203 | £28,099 | £25,358 | £18,523 | £32,193 | £50,253 | £29,179 | £71,326 | <0.05 | 1 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| technology vs. comparator in base case | | | | | | | | | | | |
| More than one CUA submitted | 12% | 7% | 16% | 17% | 12% | 21% | 6% | 3% | 10% | <0.01 | 2 |
| If More than one CUA submitted - low range | £16,256 | £4,175 | £28,336 | £13,141 | £9,888 | £16,393 | £20,201 | £10,459 | £29,944 | <0.01 | 3 |
| If More than one CUA submitted - high range | £95,652 | £6,764 | £184,540 | £125,859 | £82,185 | £169,533 | £69,600 | £28,140 | £111,059 | <0.01 | 3 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | £11,278 | £6,677 | £15,879 | £14,881 | £6,640 | £23,123 | £59,904 | £17,958 | £101,849 | <0.05 | 1 |
| Uncertainty around base case ICER reported in submission (univariate) Low | £105,042 | £29,268 | £180,816 | £42,049 | £28,653 | £55,445 | £242,362 | £75,624 | £409,101 | <0.05 | 1 |
| Uncertainty around base case ICER reported in submission (univariate) High | 65% | 51% | 78% | 46% | 36% | 57% | 23% | 11% | 35% | <0.05 | 1 |
| Non-CUA analyses submitted | 23% | 17% | 28% | 21% | 16% | 26% | 26% | 20% | 33% | NS | 2 |
| Anticipated budgetary impact of introduction of new technology in health care system | £10.9 | £6 | £15.3 | £274.6 | £25.4 | £523.8 | £163.4 | -£40 | £366.7 | <0.05 | 1 |
| Prevalence of disease/clinical condition | 105,118 | 44,430 | 165,805 | 746,591 | 363,878 | 1,129,304 | 561,944 | 326,147 | 797,741 | <0.05 | 1 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Societal Perspective adopted | 1% | 0% | 3% | 1% | 0% | 3% | 0% | 0% | 0% | NS | 2 |
| Availability of alternative therapies in current treatment setting. | 82% | 77% | 86% | 86% | 83% | 90% | 85% | 81% | 89% | NS | 2 |
| Inclusion of patient submission | 17% | 12% | 22% | 27% | 22% | 32% | 22% | 18% | 27% | <0.05 | 2 |
| Number of Decision Makers Accountable | 24 | 23 | 25 | 27 | 26 | 27 | 27 | 27 | 28 | <0.01 | 1 |
| Cost-effectiveness evaluation component in process | 64% | 58% | 70% | 60% | 55% | 65% | 54% | 49% | 60% | <0.05 | 2 |
| Budget impact as a component of decision-making process | 81% | 76% | 86% | 68% | 63% | 73% | 57% | 52% | 62% | <0.01 | 2 |
| Pricing known during appraisal | 49% | 43% | 55% | 23% | 18% | 27% | 12% | 9% | 16% | <0.01 | 2 |
| Number of drugs appraised in same appraisal | 1.1 | 1.1 | 1.2 | 1.4 | 1.3 | 1.6 | 1.1 | 1.0 | 1.1 | <0.05 | 3 |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | 2 |
| Independence of decision-making agency | 80% | 75% | 85% | 73% | 69% | 78% | 60% | 55% | 65% | <0.01 | 2 |
| Date guidance was issued | 2006 | 2006 | 2006 | 2006 | 2,006 | 2007 | 2007 | 2007 | 2007 | <0.01 | 1 |
| Population size – Agency coverage (million) | 27.50 | 24.80 | 30.20 | 35.30 | 32.70 | 37.90 | 34.10 | 31.10 | 37.10 | <0.01 | 1 |
| GDP-healthcare expenditure | 10% | 9% | 10% | 10% | 9% | 10% | 10% | 10% | 10% | <0.01 | 3 |
| Healthcare expenditure on | £263 | £252 | £274 | £280 | £269 | £291 | £306 | £293 | £320 | <0.01 | 1 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pharmaceuticals | | | | | | | | | | | |
| Election year at time of decision | 28% | 22% | 33% | 31% | 26% | 35% | 30% | 25% | 35% | NS | 2 |
| Priority disease area | 64% | 58% | 70% | 62% | 57% | 67% | 63% | 58% | 68% | NS | 2 |
| Orphan Designated | 11% | 7% | 15% | 9% | 6% | 12% | 8% | 5% | 11% | NS | 2 |
| cardiovascular system | 10% | 6% | 13% | 12% | 9% | 15% | 11% | 7% | 14% | NS | 2 |
| central nervous system | 12% | 8% | 15% | 19% | 15% | 23% | 23% | 19% | 28% | <0.01 | 2 |
| ear, nose and oropharynx | 1% | 0% | 3% | 0% | 0% | 0% | 0% | 0% | 1% | <0.05 | 2 |
| endocrine system | 4% | 2% | 7% | 7% | 5% | 10% | 6% | 4% | 9% | NS | 2 |
| eye | 3% | 1% | 5% | 1% | 0% | 2% | 1% | 0% | 2% | <0.05 | 2 |
| gastro-intestinal system | 3% | 1% | 6% | 2% | 1% | 4% | 6% | 3% | 9% | <0.05 | 2 |
| infections | 9% | 6% | 12% | 12% | 9% | 16% | 10% | 7% | 13% | NS | 2 |
| malignant disease and immunosuppression | 34% | 29% | 40% | 22% | 18% | 26% | 21% | 17% | 26% | <0.01 | 2 |
| musculoskeletal and joint diseases | 10% | 7% | 14% | 12% | 9% | 15% | 5% | 2% | 7% | <0.01 | 2 |
| nutrition and blood | 6% | 3% | 9% | 4% | 2% | 6% | 5% | 3% | 8% | NS | 2 |
| obstetrics, gynaecology, and urinary-tract disorders | 0% | 0% | 1% | 2% | 1% | 4% | 1% | 0% | 2% | <0.05 | 2 |
| respiratory system | 1% | 0% | 2% | 1% | 0% | 2% | 6% | 3% | 9% | <0.01 | 2 |
| skin | 4% | 2% | 7% | 5% | 3% | 7% | 4% | 2% | 6% | NS | 2 |

**Table 2. Descriptive Statistics of pooled sample of coverage decisions from NICE, SMC, CFH and HAS (n=977)**

| | NICE Total (n=118) | | | SMC Total (n=288) | | | CFH Total (n=256) | | | HAS Total (n=315) | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | mean | 95% CI | | | |
| Number of RCTs considered in decision | 6.7 | 5.2 | 8.4 | 2.2 | 1.9 | 2.5 | 2.6 | 2.3 | 3.0 | 2.3 | 2.1 | 2.6 | <0.01 | 1 |
| Size of population included in RCTs | 1249 | 807 | 1691 | 991 | 689 | 1294 | 830 | 494 | 1165 | 1154 | 824 | 1484 | <0.01 | 3 |
| Statistically Significant results - yes | 39% | 29% | 47% | 58% | 48% | 60% | 20% | 33% | 45% | 47% | 33% | 43% | <0.01 | 2 |
| no | 16% | 9% | 22% | 19% | 13% | 22% | 46% | 13% | 22% | 20% | 12% | 20% | <0.01 | 2 |
| inconsistent | 45% | 34% | 52% | 23% | 16% | 26% | 34% | 23% | 34% | 33% | 21% | 31% | NS | 2 |
| Length/extent of follow-up in RCT | 76.2 | 63.5 | 88.9 | 44.9 | 38.4 | 51.4 | 39.5 | 33.4 | 45.6 | 49.3 | 41.2 | 57.4 | <0.01 | 1 |
| Relevance of RCT to payor decision | 47% | 39% | 55% | 52% | 46% | 57% | 44% | 38% | 51% | 42% | 37% | 48% | <0.01 | 3 |
| Number of observational studies considered in guidance | 0.6 | 0.1 | 1.1 | 1.3 | -0.4 | 2.9 | 0.6 | 0.4 | 0.7 | 0.3 | 0.2 | 0.4 | <0.01 | 1 |
| Consideration of Cost Utility Analysis in guidance | 95% | 91% | 99% | 74% | 69% | 79% | 11% | 7% | 15% | 0% | 0% | 0% | <0.01 | 2 |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | £31,266 | £29,122 | £42,410 | £34,055 | £21,630 | £46,481 | £30,977 | £8,643 | £53,312 | | | | <0.01 | 3 |
| More than one CUA submitted | 63% | 54% | 72% | 1% | 0% | 2% | 1% | 0% | 2% | | | | <0.01 | 1 |
| If More than one CUA submitted - low range | £13,260 | £10,409 | £16,110 | £10,399 | -£9,782 | £30,580 | £85,091 | -£198,747 | £368,928 | | | | <0.10 | 3 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| If More than one CUA submitted - high range | £107,421 | £66,886 | £147,956 | £18,207 | -£2,672 | £39,086 | £221,499 | £201,610 | £241,389 | | | | <0.05 | 3 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | 43% | 34% | 52% | 57% | 42% | 73% | 66% | 42% | 90% | | | | <0.10 | 3 |
| Uncertainty around base case ICER reported in submission (univariate) Low | £25,417 | £6,412 | £44,422 | £33,277 | £8,916 | £57,637 | £15,187 | £4,338 | £26,036 | | | | <0.05 | 3 |
| Uncertainty around base case ICER reported in submission (univariate) High | £167,389 | £56,865 | £277,913 | £77,927 | £19,847 | £43,448 | £92,826 | -£6,356 | £192,008 | | | | <0.01 | 3 |
| Non-CUA analyses submitted | 23% | 15% | 30% | 30% | 25% | 36% | 15% | 11% | 20% | | | | <0.01 | 2 |
| Anticipated budgetary impact of introduction of new technology in health care system | £701.3 | £179.7 | £1,223.0 | £1.2 | £0.9 | £1.5 | £31.0 | £5.9 | £56.2 | | | | | |
| Prevalence of disease/clinical condition | 2,418,119 | 1,256,514 | 3,579,724 | 11,277 | 1,647 | 20,908 | 94,543 | 31,394 | 157,693 | 511,047 | 314,122 | 707,972 | <0.01 | 1 |
| Social Perspective adopted | 3% | 0% | 5% | 0% | 0% | 1% | 1% | 0% | 1% | 0% | 0% | 0% | <0.10 | 2 |
| Availability of alternative therapies in current treatment setting. | 89% | 83% | 95% | 83% | 79% | 88% | 79% | 74% | 84% | 89% | 85% | 92% | <0.01 | 2 |
| Inclusion of patient submission | 87% | 81% | 93% | 42% | 36% | 48% | 4% | 2% | 6% | 0% | 0% | 0% | <0.01 | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Decision Makers Accountable | 30 | 28 | 32 | 25 | 24 | 25 | 20 | 20 | 20 | 31 | 31 | 31 | <0.01 | 1 |
| Cost-effectiveness evaluation component in process | 100% | 100% | 100% | 100% | 100% | 100% | 67% | 61% | 73% | 0% | 0% | 0% | <0.01 | 2 |
| Budget impact as a component of decision-making process | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | <0.01 | 2 |
| Pricing and Reimbursement decided jointly | 0% | 0% | 0% | 0% | 0% | 0% | 100% | 100% | 100% | 0% | 0% | 0% | <0.01 | 2 |
| Number of drugs appraised in same appraisal | 2.8 | 2.5 | 3.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | <0.01 | 1 |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | 2 |
| Independence of decision-making agency | 100% | 100% | 100% | 0% | 0% | 0^% | 100% | 100% | 100% | 100% | 100% | 100% | <0.01 | 2 |
| Date guidance was issued | 2,007 | 2,006 | 2,007 | 2,006 | 2,006 | 2006 | 2006 | 2006 | 2006 | 2007 | 2006 | 2007 | <0.01 | 1 |
| Population size – Agency coverage | 53.90 | 53.80 | 54.00 | 5.13 | 5.12 | 5.13 | 16.30 | 16.30 | 16.40 | 63.40 | 63.30 | 63.50 | <0.01 | 3 |
| GDP-healthcare expenditure | 8% | 8% | 8% | 8% | 8% | 8% | 10% | 10% | 10% | 11% | 11% | 11% | <0.01 | 1 |
| Healthcare expenditure on pharmaceuticals | £175 | £173 | £176 | £190 | £189 | £190 | £249 | £246 | £251 | £439 | £437 | £441 | <0.01 | 3 |
| Election year at time of decision | 7% | 2% | 11% | 40% | 35% | 45% | 20% | 15% | 24% | 30% | 25% | 36% | <0.01 | 2 |
| Priority disease area | 56% | 47% | 65% | 66% | 61% | 71% | 55% | 49% | 61% | 70% | 65% | 75% | <0.01 | 2 |
| Orphan Designated | 3% | 0% | 5% | 11% | 8% | 15% | 9% | 5% | 12% | 9% | 6% | 12% | <0.05 | 2 |
| cardiovascular system | 10% | 5% | 16% | 11% | 7% | 14% | 9% | 5% | 12% | 14% | 10% | 18% | NS | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| central nervous system | 15% | 9% | 22% | 22% | 16% | 26% | 16% | 12% | 21% | 18% | 14% | 23% | NS | 2 |
| ear, nose and oropharynx | 0% | 0% | 0% | 0% | 0% | 1% | 1% | 0% | 2% | 0% | 0% | 1% | NS | 2 |
| endocrine system | 1% | -1% | 3% | 9% | 5% | 12% | 6% | 3% | 9% | 6% | 4% | 9% | <0.05 | 2 |
| eye | 2% | -1% | 4% | 1% | 0% | 2% | 2% | 0% | 3% | 2% | 0% | 3% | NS | 2 |
| gastro-intestinal system | 2% | -1% | 4% | 4% | 2% | 6% | 5% | 2% | 7% | 3% | 1% | 6% | NS | 2 |
| infections | 12% | 6% | 18% | 10% | 7% | 14% | 9% | 5% | 12% | 12% | 8% | 15% | NS | 2 |
| malignant disease and immunosuppression | 31% | 22% | 39% | 23% | 20% | 30% | 25% | 19% | 30% | 24% | 19% | 29% | NS | 2 |
| musculoskeletal and joint diseases | 19% | 12% | 27% | 3% | 1% | 6% | 10% | 6% | 14% | 9% | 6% | 12% | <0.01 | 2 |
| nutrition and blood | 3% | 0% | 5% | 6% | 3% | 9% | 7% | 4% | 10% | 3% | 1% | 6% | NS | 2 |
| obstetrics, gynaecology, and urinary-tract disorders | 0% | 0% | 0% | 2% | 0% | 3% | 3% | 1% | 5% | 0% | 0% | 1% | <0.05 | 2 |
| respiratory system | 1% | -1% | 3% | 3% | 2% | 6% | 4% | 1% | 6% | 2% | 1% | 4% | NS | 2 |
| skin | 5% | 1% | 9% | 3% | 2% | 6% | 5% | 2% | 7% | 4% | 2% | 7% | NS | 2 |

Note: 1= Both ANOVA and Kruskal-Wallis test both indicate similar level of statistical significance. 2=Chi-squared test used, as categorical variable.

3 = Either ANOVA or Kruskal-Wallis test indicate statistical significance.

**Table 3. Descriptive Statistics of pooled sample of coverage decisions from NICE, SMC, CFH and HAS (n= 267) for Recommended Technologies**

| | NICE | | | SMC | | | CFH | | | HAS | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | | |
| Number of RCTs considered in decision | 6.8 | 2.1 | 11.4 | 2.3 | 1.7 | 2.9 | 2.2 | 1.8 | 2.5 | 1.7 | 1.3 | 2.2 | <0.01 | 1 |
| Size of population included in RCTs | 1765 | 804 | 2727 | 1532 | 572 | 2492 | 525 | 366 | 684 | 836 | 192 | 1480 | <0.01 | 3 |
| Statistically Significant results - yes | 59% | 41% | 77% | 48% | 34% | 62% | 45% | 37% | 54% | 53% | 39% | 67% | NS | 2 |
| Inconsistent | 6% | -3% | 15% | 24% | 12% | 36% | 18% | 11% | 24% | 6% | -1% | 13% | NS | 2 |
| No | 31% | 14% | 48% | 22% | 11% | 34% | 20% | 13% | 27% | 10% | 1% | 18% | <0.05 | 2 |
| Length/extent of follow-up in RCT | 95.9 | 65.9 | 125.8 | 59.7 | 41.7 | 77.7 | 44.0 | 34.5 | 53.5 | 67.7 | 50.6 | 84.8 | <0.01 | 3 |
| Use of active comparator | 63% | 46% | 80% | 67% | 55% | 79% | 51% | 42% | 60% | 25% | 11% | 38% | <0.01 | 1 |
| Number of observational studies considered in guidance | 1.5 | -0.4 | 3.4 | 0.0 | 0.0 | 0.0 | 0.6 | 0.3 | 0.8 | 0.6 | 0.0 | 1.2 | <0.05 | 1 |
| Consideration of Cost Utility Analysis in guidance | 100% | 100% | 100% | 67% | 54% | 80% | 12% | 7% | 18% | 0% | 0% | 0% | <0.01 | 2 |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | £17,782 | £11,066 | £24,498 | £11,893 | £8,645 | £15,140 | £43,864 | £243,619 | £673 | . | . | . | NS | 1 |
| More than one CUA | 69% | 52% | 86% | 0% | 0% | 0% | 2% | 0% | 5% | . | . | . | <0.01 | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| submitted | | | | | | | | | | | | | | |
| If More than one CUA submitted - low range | £8,607 | £4,551 | £12,664 | . | . | . | £85,091 | £107,429 | £62,752 | . | . | . | <0.05 | 3 |
| If More than one CUA submitted - high range | £83,666 | -£12,619 | £179,951 | . | . | . | £221,499 | £223,064 | £219,934 | . | . | . | <0.05 | 3 |
| Uncertainty around base case ICER reported in submission (univariate) Low | £7,881 | £5,234 | £10,527 | £11,251 | £5,017 | £17,486 | £26,614 | £62,752 | £5,075 | . | . | . | NS | 3 |
| Uncertainty around base case ICER reported in submission (univariate) High | £113,286 | £1,293 | £225,279 | £31,647 | £19,847 | £43,448 | £163,012 | £778,584 | £11,639 | . | . | . | NS | 1 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | 61% | 45% | 77% | 72% | -164% | 307% | 79% | 43% | 115% | . | . | . | NS | 1 |
| Non-CUA analyses submitted | 22% | 7% | 37% | 39% | 25% | 52% | 16% | 10% | 23% | . | . | . | <0.01 | 2 |
| Anticipated budgetary impact of introduction of new technology in health care system (million) | £35.9 | £17.9 | £53.9 | £1.9 | £1.0 | £2.8 | £5.8 | £38.1 | £0.0 | . | . | . | <0.01 | 1 |
| Prevalence of disease/clinical condition | 392,063 | 82,017 | 702,109 | 36,122 | -12,168 | 84,412 | 61,816 | -16,081 | 139,713 | 68,281 | -36,271 | 172,834 | <0.01 | 1 |
| Societal Perspective adopted | 3% | -3% | 9% | 0% | 0% | 0% | 1% | -1% | 3% | . | . | . | NS | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Availability of alternative therapies in current treatment setting. | 88% | 75% | 100% | 93% | 85% | 100% | 76% | 68% | 83% | 81% | 69% | 93% | <0.05 | 2 |
| Inclusion of patient submission | 91% | 80% | 101% | 27% | 14% | 41% | 2% | 0% | 5% | 0% | 0% | 0% | <0.01 | 2 |
| Number of Decision Makers Accountable | 28 | 25 | 31 | 25 | 24 | 26 | 20 | 20 | 20 | 31 | 31 | 31 | <0.01 | 1 |
| Cost-effectiveness evaluation component in process | 100% | 100% | 100% | 100% | 100% | 100% | 67% | 61% | 73% | 0% | 0% | 0% | <0.01 | 2 |
| Budget impact as a component of decision-making process | 100% | 100% | 100% | 100% | 100% | 100% | 65% | 57% | 74% | 0% | 0% | 0% | <0.01 | 2 |
| Pricing known during appraisal | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | <0.01 | 2 |
| Number of drugs appraised in same appraisal | 2.0 | 1.5 | 2.5 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | <0.01 | 1 |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | 2 |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | 100% | 100% | 100% | <0.01 | 2 |
| Date guidance was issued | 2,006 | 2,006 | 2,007 | 2,007 | 2,006 | 2007 | 2006 | 2006 | 2006 | 2006 | 2006 | 2006 | <0.05 | 3 |
| Population size – Agency coverage | 53.80 | 53.70 | 54.00 | 5.14 | 5.13 | 5.15 | 16.30 | 16.30 | 16.40 | 63.20 | 63.00 | 63.40 | <0.01 | 3 |
| GDP-healthcare expenditure | 8% | 8% | 8% | 8% | 8% | 8% | 10% | 10% | 10% | 11% | 11% | 11% | <0.01 | 1 |
| Healthcare expenditure on pharmaceuticals | £173 | £171 | £175 | £191 | £190 | £193 | £293 | £289 | £296 | £209 | £232 | £188 | <0.01 | 3 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Election year at time of decision | 6% | -3% | 15% | 48% | 34% | 62% | 23% | 16% | 30% | 31% | 18% | 45% | <0.01 | 2 |
| Priority disease area | 59% | 41% | 77% | 67% | 54% | 80% | 61% | 52% | 69% | 73% | 60% | 85% | NS | 2 |
| Orphan Designated | 3% | -3% | 9% | 6% | -1% | 12% | 11% | 5% | 16% | 24% | 11% | 36% | <0.01 | 2 |
| cardiovascular system | 13% | 0% | 25% | 17% | 6% | 27% | 4% | 0% | 7% | 16% | 5% | 26% | <0.05 | 2 |
| central nervous system | 6% | -3% | 15% | 15% | 5% | 25% | 15% | 9% | 22% | 2% | -2% | 6% | <0.05 | 2 |
| ear, nose and oropharynx | 0% | 0% | 0% | 2% | -2% | 6% | 2% | 0% | 5% | 0% | 0% | 0% | NS | 2 |
| endocrine system | 0% | 0% | 0% | 9% | 1% | 17% | 5% | 1% | 8% | 2% | -2% | 6% | NS | 2 |
| eye | 3% | -3% | 9% | 4% | -1% | 9% | 3% | 0% | 6% | 2% | -2% | 6% | NS | 2 |
| gastro-intestinal system | 0% | 0% | 0% | 2% | -2% | 6% | 5% | 1% | 9% | 2% | -2% | 6% | NS | 2 |
| infections | 19% | 4% | 33% | 17% | 6% | 27% | 6% | 2% | 10% | 2% | -2% | 6% | <0.01 | 2 |
| malignant disease and immunosuppression | 41% | 23% | 59% | 20% | 9% | 31% | 37% | 29% | 45% | 39% | 25% | 53% | NS | 2 |
| musculoskeletal and joint diseases | 16% | 2% | 29% | 0% | 0% | 0% | 7% | 3% | 11% | 27% | 15% | 40% | <0.01 | 2 |
| nutrition and blood | 0% | 0% | 0% | 6% | -1% | 12% | 8% | 3% | 12% | 8% | 0% | 15% | NS | 2 |
| obstetrics, gynaecology, and urinary-tract disorders | 0% | 0% | 0% | 0% | 0% | 0% | 1% | -1% | 2% | 0% | 0% | 0% | NS | 2 |
| respiratory system | 0% | 0% | 0% | 2% | -2% | 6% | 2% | -1% | 4% | 0% | 0% | 0% | NS | 2 |
| skin | 3% | -3% | 9% | 7% | 0% | 15% | 5% | 1% | 9% | 0% | 0% | 0% | NS | 2 |

Note: 1= Both ANOVA and Kruskal-Wallis test both indicate similar level of statistical significance. 2=Chi-squared test used, as categorical variable.

3 = Either ANOVA or Kruskal-Wallis test indicate statistical significance.

**Table 4. Descriptive Statistics of pooled sample of coverage decisions from NICE, SMC, CFH and HAS (n=379 ) for Restricted Technologies**

| | NICE recommended technologies | | | SMC recommended technologies | | | CFH recommended technologies | | | HAS Recommended Technologies | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | | |
| Number of RCTs considered in decision | 7.6 | 5.9 | 9.2 | 2.3 | 1.7 | 2.8 | 2.8 | 2.2 | 3.5 | 2.3 | 1.9 | 2.7 | <0.01 | 1 |
| Size of population included in RCTs | 1044 | 442 | 1645 | 1360 | 691 | 2028 | 1364 | 447 | 2281 | 1206 | 663 | 1750 | NS | 1 |
| Statistically Significant results - yes | 29% | 18% | 40% | 52% | 42% | 62% | 35% | 25% | 45% | 47% | 38% | 56% | <0.01 | 2 |
| Inconsistent | 17% | 8% | 27% | 22% | 13% | 30% | 17% | 9% | 26% | 11% | 5% | 16% | <0.01 | 2 |
| no | 51% | 39% | 63% | 23% | 14% | 31% | 35% | 25% | 45% | 27% | 19% | 35% | NS | 2 |
| Length/extent of follow-up in RCT | 65.8 | 51.5 | 80.2 | 50.5 | 37.6 | 63.3 | 30.9 | 23.8 | 38.0 | 50.2 | 37.4 | 63.0 | <0.01 | 3 |
| Use of active comparator | 40% | 30% | 50% | 53% | 44% | 62% | 45% | 35% | 55% | 40% | 31% | 48% | NS | 1 |
| Number of observational studies considered in guidance | 0.3 | 0.1 | 0.5 | 0.1 | 0.0 | 0.2 | 0.6 | 0.2 | 0.9 | 0.3 | 0.1 | 0.6 | <0.05 | 3 |
| Consideration of Cost Utility Analysis in guidance | 94% | 89% | 100% | 77% | 69% | 86% | 7% | 1% | 12% | 0% | 0% | 0% | <0.01 | 2 |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | £24,867 | £21,002 | £28,731 | £26,316 | £13,265 | £39,367 | £15,195 | £9,592 | £20,799 | . | . | . | <0.01 | 3 |
| More than one CUA submitted | 63% | 51% | 75% | 1% | -1% | 3% | 0% | 0% | 0% | . | . | . | <0.01 | 2 |
| If More than one CUA submitted - | £13,417 | £10,119 | £16,714 | £3,194 | . | . | . | . | . | . | . | . | NS | 1 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| low range | | | | | | | | | | | | | | |
| If More than one CUA submitted - high range | £129,690 | £85,202 | £174,179 | £10,920 | . | . | . | . | . | . | . | . | <0.05 | 3 |
| Uncertainty around base case ICER reported in submission (univariate) Low | £19,747 | £4,460 | £35,035 | £8,963 | £5,794 | £12,132 | £11,336 | -£14,217 | £36,890 | . | . | . | NS | 1 |
| Uncertainty around base case ICER reported in submission (univariate) High | £57,146 | £29,962 | £84,329 | £27,112 | £20,849 | £33,375 | £40,715 | -£36,536 | £117,965 | . | . | . | <0.01 | 3 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | 41% | 29% | 53% | 71% | 53% | 90% | 52% | -66% | 170% | . | . | . | NS | 1 |
| Non-CUA analyses submitted | 28% | 17% | 38% | 25% | 16% | 33% | 10% | 4% | 17% | . | . | . | <0.05 | 2 |
| Anticipated budgetary impact of introduction of new technology in health care system | £828.5 | £51.3 | £1,605.6 | £1.3 | £0.8 | £1.8 | £66.8 | -£18 | £151.1 | . | . | . | <0.01 | 1 |
| Prevalence of disease/clinical condition | 3,194,243 | 1,419,388 | 4,969,097 | 6,584 | 1,612 | 11,557 | 41,087 | 8,932 | 73,242 | 236,387 | 113,270 | 359,503 | <0.01 | 1 |
| Societal Perspective adopted | 3% | -1% | 7% | 1% | -1% | 3% | 0% | 0% | 0% | . | . | . | NS | 2 |
| Availability of alternative therapies in current treatment setting. | 90% | 83% | 97% | 85% | 78% | 92% | 85% | 77% | 93% | 85% | 79% | 92% | NS | 2 |
| Inclusion of patient | 84% | 75% | 93% | 45% | 34% | 56% | 2% | -1% | 6% | 0% | 0% | 0% | <0.01 | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| submission | | | | | | | | | | | | | | |
| Number of Decision Makers Accountable | 31 | 29 | 34 | 25 | 24 | 25 | 20 | 20 | 20 | 31 | 31 | 31 | <0.01 | 1 |
| Cost-effectiveness evaluation component in process | 100% | 100% | 100% | 100% | 100% | 100% | 64% | 54% | 74% | 0% | 0% | 0% | <0.01 | 2 |
| Budget impact as a component of decision-making process | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | <0.01 | 2 |
| Pricing known during appraisal | 0% | 0% | 0% | 0% | 0% | 0% | 100% | 100% | 100% | 0% | 0% | 0% | <0.01 | 2 |
| Number of drugs appraised in same appraisal | 3.4 | 2.9 | 3.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | <0.01 | 1 |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | 2 |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | 100% | 100% | 100% | <0.01 | 2 |
| Date guidance was issued | 2007 | 2006 | 2007 | 2007 | 2,006 | 2007 | 2006 | 2006 | 2006 | 2006 | 2006 | 2007 | <0.01 | 3 |
| Population size – Agency coverage (million) | 53.90 | 53.80 | 54.10 | 5.13 | 5.13 | 5.14 | 16.30 | 16.30 | 16.30 | 63.30 | 63.20 | 63.40 | <0.01 | 3 |
| GDP-healthcare expenditure | 8% | 8% | 8% | 8% | 8% | 8% | 10% | 10% | 10% | 11% | 11% | 11% | <0.01 | 1 |
| Healthcare expenditure on pharmaceuticals | £174 | £173 | £176 | £191 | £190 | £192 | £247 | £243 | £251 | £437 | £434 | £440 | <0.01 | 3 |
| Election year at time of decision | 7% | 1% | 14% | 48% | 38% | 58% | 20% | 11% | 28% | 37% | 28% | 46% | <0.01 | 2 |
| Priority disease area | 57% | 45% | 69% | 69% | 59% | 78% | 43% | 32% | 54% | 74% | 66% | 82% | <0.01 | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Orphan Designated | 3% | -1% | 7% | 9% | 3% | 14% | 9% | 3% | 16% | 12% | 6% | 18% | NS | 2 |
| cardiovascular system | 12% | 4% | 19% | 9% | 3% | 14% | 15% | 7% | 23% | 13% | 7% | 19% | NS | 2 |
| central nervous system | 22% | 12% | 32% | 19% | 11% | 26% | 15% | 7% | 23% | 20% | 13% | 27% | NS | 2 |
| ear, nose and oropharynx | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | 2 |
| endocrine system | 1% | -1% | 4% | 10% | 4% | 16% | 8% | 2% | 14% | 7% | 3% | 12% | NS | 2 |
| eye | 0% | 0% | 0% | 1% | -1% | 3% | 0% | 0% | 0% | 2% | -1% | 4% | NS | 2 |
| gastro-intestinal system | 1% | -1% | 4% | 1% | -1% | 3% | 5% | 0% | 9% | 2% | -1% | 4% | NS | 2 |
| infections | 7% | 1% | 14% | 16% | 9% | 23% | 12% | 5% | 19% | 13% | 7% | 19% | NS | 2 |
| malignant disease and immunosuppression | 23% | 13% | 33% | 26% | 18% | 35% | 9% | 3% | 16% | 27% | 19% | 35% | <0.05 | 2 |
| musculoskeletal and joint diseases | 22% | 12% | 32% | 5% | 1% | 9% | 19% | 10% | 27% | 7% | 3% | 12% | <0.01 | 2 |
| nutrition and blood | 4% | -1% | 9% | 3% | 0% | 6% | 8% | 2% | 14% | 2% | 0% | 5% | NS | 2 |
| obstetrics, gynaecology, and urinary-tract disorders | 0% | 0% | 0% | 3% | 0% | 6% | 7% | 1% | 12% | 0% | 0% | 0% | <0.01 | 2 |
| respiratory system | 1% | -1% | 4% | 3% | 0% | 6% | 0% | 0% | 0% | 1% | -1% | 2% | NS | 2 |
| skin | 6% | 0% | 11% | 5% | 1% | 9% | 2% | -1% | 6% | 6% | 2% | 10% | NS | 2 |

Note: 1= Both ANOVA and Kruskal-Wallis test both indicate similar level of statistical significance. 2=Chi-squared test used, as categorical variable.

3 = Either ANOVA or Kruskal-Wallis test indicate statistical significance.

**Table 5. Descriptive Statistics of pooled sample of coverage decisions from NICE, SMC, CFH and HAS (n= 267) for Recommended Technologies**

| | NICE | | | SMC | | | CFH | | | HAS | | | P value | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI | | | |
| Number of RCTs considered in decision | 6.8 | 2.1 | 11.4 | 2.3 | 1.7 | 2.9 | 2.2 | 1.8 | 2.5 | 1.7 | 1.3 | 2.2 | <0.01 | 1 |
| Size of population included in RCTs | 1765 | 804 | 2727 | 1532 | 572 | 2492 | 525 | 366 | 684 | 836 | 192 | 1480 | <0.01 | 3 |
| Statistically Significant results - yes | 59% | 41% | 77% | 48% | 34% | 62% | 45% | 37% | 54% | 53% | 39% | 67% | NS | 2 |
| Inconsistent | 6% | -3% | 15% | 24% | 12% | 36% | 18% | 11% | 24% | 6% | -1% | 13% | NS | 2 |
| No | 31% | 14% | 48% | 22% | 11% | 34% | 20% | 13% | 27% | 10% | 1% | 18% | <0.05 | 2 |
| Length/extent of follow-up in RCT | 95.9 | 65.9 | 125.8 | 59.7 | 41.7 | 77.7 | 44.0 | 34.5 | 53.5 | 67.7 | 50.6 | 84.8 | <0.01 | 3 |
| Use of active comparator | 63% | 46% | 80% | 67% | 55% | 79% | 51% | 42% | 60% | 25% | 11% | 38% | <0.01 | 1 |
| Number of observational studies considered in guidance | 1.5 | -0.4 | 3.4 | 0.0 | 0.0 | 0.0 | 0.6 | 0.3 | 0.8 | 0.6 | 0.0 | 1.2 | <0.05 | 1 |
| Consideration of Cost Utility Analysis in guidance | 100% | 100% | 100% | 67% | 54% | 80% | 12% | 7% | 18% | 0% | 0% | 0% | <0.01 | 2 |
| Incremental Cost-effectiveness ratio of technology vs. comparator in base case | £17,782 | £11,066 | £24,498 | £11,893 | £8,645 | £15,140 | £43,864 | £243,619 | £673 | . | . | . | NS | 1 |
| More than one CUA | 69% | 52% | 86% | 0% | 0% | 0% | 2% | 0% | 5% | . | . | . | <0.01 | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| submitted | | | | | | | | | | | | | | |
| If More than one CUA submitted - low range | £8,607 | £4,551 | £12,664 | . | . | . | £85,091 | £107,429 | £62,752 | . | . | . | <0.05 | 3 |
| If More than one CUA submitted - high range | £83,666 | -£12,619 | £179,951 | . | . | . | £221,499 | £223,064 | £219,934 | . | . | . | <0.05 | 3 |
| Uncertainty around base case ICER reported in submission (univariate) Low | £7,881 | £5,234 | £10,527 | £11,251 | £5,017 | £17,486 | £26,614 | £62,752 | £5,075 | . | . | . | NS | 3 |
| Uncertainty around base case ICER reported in submission (univariate) High | £113,286 | £1,293 | £225,279 | £31,647 | £19,847 | £43,448 | £163,012 | £778,584 | £11,639 | . | . | . | NS | 1 |
| Uncertainty around the base case ICER reported in submission (probabilistic) | 61% | 45% | 77% | 72% | -164% | 307% | 79% | 43% | 115% | . | . | . | NS | 1 |
| Non-CUA analyses submitted | 22% | 7% | 37% | 39% | 25% | 52% | 16% | 10% | 23% | . | . | . | <0.01 | 2 |
| Anticipated budgetary impact of introduction of new technology in health care system (million) | £35.9 | £17.9 | £53.9 | £1.9 | £1.0 | £2.8 | £5.8 | £38.1 | £0.0 | . | . | . | <0.01 | 1 |
| Prevalence of disease/clinical condition | 392,063 | 82,017 | 702,109 | 36,122 | -12,168 | 84,412 | 61,816 | -16,081 | 139,713 | 68,281 | -36,271 | 172,834 | <0.01 | 1 |
| Societal Perspective adopted | 3% | -3% | 9% | 0% | 0% | 0% | 1% | -1% | 3% | . | . | . | NS | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Availability of alternative therapies in current treatment setting. | 88% | 75% | 100% | 93% | 85% | 100% | 76% | 68% | 83% | 81% | 69% | 93% | <0.05 | 2 |
| Inclusion of patient submission | 91% | 80% | 101% | 27% | 14% | 41% | 2% | 0% | 5% | 0% | 0% | 0% | <0.01 | 2 |
| Number of Decision Makers Accountable | 28 | 25 | 31 | 25 | 24 | 26 | 20 | 20 | 20 | 31 | 31 | 31 | <0.01 | 1 |
| Cost-effectiveness evaluation component in process | 100% | 100% | 100% | 100% | 100% | 100% | 67% | 61% | 73% | 0% | 0% | 0% | <0.01 | 2 |
| Budget impact as a component of decision-making process | 100% | 100% | 100% | 100% | 100% | 100% | 65% | 57% | 74% | 0% | 0% | 0% | <0.01 | 2 |
| Pricing known during appraisal | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | <0.01 | 2 |
| Number of drugs appraised in same appraisal | 2.0 | 1.5 | 2.5 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | <0.01 | 1 |
| Accountability of drug budget | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | n/a | 2 |
| Independence of decision-making agency | 100% | 100% | 100% | 100% | 100% | 100% | 0% | 0% | 0% | 100% | 100% | 100% | <0.01 | 2 |
| Date guidance was issued | 2,006 | 2,006 | 2,007 | 2,007 | 2,006 | 2007 | 2006 | 2006 | 2006 | 2006 | 2006 | 2006 | <0.05 | 3 |
| Population size – Agency coverage | 53.80 | 53.70 | 54.00 | 5.14 | 5.13 | 5.15 | 16.30 | 16.30 | 16.40 | 63.20 | 63.00 | 63.40 | <0.01 | 3 |
| GDP-healthcare expenditure | 8% | 8% | 8% | 8% | 8% | 8% | 10% | 10% | 10% | 11% | 11% | 11% | <0.01 | 1 |
| Healthcare expenditure on pharmaceuticals | £173 | £171 | £175 | £191 | £190 | £193 | £293 | £289 | £296 | £209 | £232 | £188 | <0.01 | 3 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Election year at time of decision | 6% | -3% | 15% | 48% | 34% | 62% | 23% | 16% | 30% | 31% | 18% | 45% | <0.01 | 2 |
| Priority disease area | 59% | 41% | 77% | 67% | 54% | 80% | 61% | 52% | 69% | 73% | 60% | 85% | NS | 2 |
| Orphan Designated | 3% | -3% | 9% | 6% | -1% | 12% | 11% | 5% | 16% | 24% | 11% | 36% | <0.01 | 2 |
| cardiovascular system | 13% | 0% | 25% | 17% | 6% | 27% | 4% | 0% | 7% | 16% | 5% | 26% | <0.05 | 2 |
| central nervous system | 6% | -3% | 15% | 15% | 5% | 25% | 15% | 9% | 22% | 2% | -2% | 6% | <0.05 | 2 |
| ear, nose and oropharynx | 0% | 0% | 0% | 2% | -2% | 6% | 2% | 0% | 5% | 0% | 0% | 0% | NS | 2 |
| endocrine system | 0% | 0% | 0% | 9% | 1% | 17% | 5% | 1% | 8% | 2% | -2% | 6% | NS | 2 |
| eye | 3% | -3% | 9% | 4% | -1% | 9% | 3% | 0% | 6% | 2% | -2% | 6% | NS | 2 |
| gastro-intestinal system | 0% | 0% | 0% | 2% | -2% | 6% | 5% | 1% | 9% | 2% | -2% | 6% | NS | 2 |
| infections | 19% | 4% | 33% | 17% | 6% | 27% | 6% | 2% | 10% | 2% | -2% | 6% | <0.01 | 2 |
| malignant disease and immunosuppression | 41% | 23% | 59% | 20% | 9% | 31% | 37% | 29% | 45% | 39% | 25% | 53% | NS | 2 |
| musculoskeletal and joint diseases | 16% | 2% | 29% | 0% | 0% | 0% | 7% | 3% | 11% | 27% | 15% | 40% | <0.01 | 2 |
| nutrition and blood | 0% | 0% | 0% | 6% | -1% | 12% | 8% | 3% | 12% | 8% | 0% | 15% | NS | 2 |
| obstetrics, gynaecology, and urinary-tract disorders | 0% | 0% | 0% | 0% | 0% | 0% | 1% | -1% | 2% | 0% | 0% | 0% | NS | 2 |
| respiratory system | 0% | 0% | 0% | 2% | -2% | 6% | 2% | -1% | 4% | 0% | 0% | 0% | NS | 2 |
| skin | 3% | -3% | 9% | 7% | 0% | 15% | 5% | 1% | 9% | 0% | 0% | 0% | NS | 2 |

Note: 1= Both ANOVA and Kruskal-Wallis test both indicate similar level of statistical significance. 2=Chi-squared test used, as categorical variable.

3 = Either ANOVA or Kruskal-Wallis test indicate statistical significance.

**Pooled analysis: preliminary Multivariate Model**

The results of the multinomial regression yield a pseudo R-squared of 0.1604, suggesting that the model explains 16% of the variability in the pooled sample of coverage decisions.

**Table 8.1 Multivariate analysis of NICE, SMC, CFH and HAS coverage decisions 2004-2009: preliminary model**

| | **Restricted vs. Recommended** | | | | **Not Recommended vs. Recommended** | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.030 | 0.192 | -0.08 | 0.02 | -0.062 | 0.062 | -0.13 | 0.00 |
| Clinical superiority demonstrated in RCT | -0.088 | 0.918 | -1.76 | 1.58 | -1.777 | 0.021 | -3.29 | -0.27 |
| Lack of clinical superiority in RCT | -0.105 | 0.904 | -1.81 | 1.60 | -1.370 | 0.083 | -2.92 | 0.18 |
| Inconsistent clinical superiority demonstrated in RCT | 0.455 | 0.604 | -1.27 | 2.17 | -0.990 | 0.213 | -2.55 | 0.57 |
| RCT duration of follow-up | -0.002 | 0.265 | -0.01 | 0.00 | -0.003 | 0.122 | -0.01 | 0.00 |
| CUA | 0.245 | 0.416 | -0.35 | 0.84 | 0.349 | 0.279 | -0.28 | 0.98 |
| ICER | 0.000 | 0.587 | 0.00 | 0.00 | 0.000 | 0.417 | 0.00 | 0.00 |
| Interaction term adjusting for no ICER within HAS decision-making | 0.000 | 0.138 | 0.00 | 0.00 | 0.000 | 0.091 | 0.00 | 0.00 |
| Use of active comparator in RCT | -0.596 | 0.008 | -1.03 | -0.16 | -0.864 | 0.000 | -1.35 | -0.38 |
| Orphan designation status | -0.117 | 0.709 | -0.73 | 0.50 | -0.856 | 0.019 | -1.57 | -0.14 |
| Patient submission included | 0.338 | 0.294 | -0.29 | 0.97 | 0.793 | 0.017 | 0.14 | 1.44 |
| Size of appraisal committee | 0.004 | 0.897 | -0.05 | 0.06 | 0.017 | 0.618 | -0.05 | 0.08 |
| Cost Effectiveness part of process | -0.291 | 0.512 | -1.16 | 0.58 | -0.438 | 0.444 | -1.56 | 0.68 |
| Budgetary assessment part of process | -4.550 | 0.058 | -9.25 | 0.15 | -4.822 | 0.065 | -9.94 | 0.30 |
| Number of technologies appraised simultaneously | 0.526 | 0.002 | 0.19 | 0.86 | 0.169 | 0.488 | -0.31 | 0.65 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Year of Appraisal | 0.116 | 0.336 | -0.12 | 0.35 | 0.196 | 0.140 | -0.06 | 0.46 |
| National population size | 0.000 | 0.113 | 0.00 | 0.00 | 0.000 | 0.003 | 0.00 | 0.00 |
| % GDP expenditure on healthcare | -66.848 | 0.288 | -190.12 | 56.42 | -135.869 | 0.058 | -276.45 | 4.71 |
| Healthcare expenditure on pharmaceuticals | 0.004 | 0.739 | -0.02 | 0.03 | 0.017 | 0.253 | -0.01 | 0.05 |
| Cardiovascular disease | 17.933 | 0.941 | -453.80 | 489.67 | -1.656 | 0.000 | -2.38 | -0.94 |
| Central nervous system | 18.008 | 0.940 | -453.75 | 489.76 | 2.302 | 0.090 | -0.36 | 4.96 |
| Endocrine system | 18.531 | 0.939 | -453.23 | 490.29 | 2.481 | 0.077 | -0.27 | 5.23 |
| Eye | 16.724 | 0.945 | -455.08 | 488.53 | 1.240 | 0.414 | -1.73 | 4.21 |
| Gastro-intestinal system | 17.709 | 0.941 | -454.03 | 489.44 | 2.678 | 0.060 | -0.11 | 5.47 |
| Infections | 18.415 | 0.939 | -453.35 | 490.19 | 2.103 | 0.128 | -0.61 | 4.81 |
| Cancer therapy | 17.434 | 0.942 | -454.28 | 489.15 | 1.444 | 0.290 | -1.23 | 4.12 |
| Musculoskeletal and joint diseases | 17.932 | 0.941 | -453.83 | 489.69 | 1.301 | 0.352 | -1.44 | 4.04 |
| Nutrition and blood | 17.806 | 0.941 | -453.96 | 489.57 | 2.115 | 0.133 | -0.64 | 4.88 |
| Obstetrics, gynaecology, and urinary-tract disorders | 20.318 | 0.933 | -451.45 | 492.09 | 4.001 | 0.024 | 0.51 | 7.49 |
| Obstetrics, gynaecology, and urinary-tract disorders | 18.033 | 0.940 | -453.71 | 489.78 | 3.441 | 0.021 | 0.52 | 6.36 |
| Skin | 18.046 | 0.940 | -453.71 | 489.80 | 2.067 | 0.145 | -0.71 | 4.85 |
| RCT follow-up not available | 0.154 | 0.743 | -0.77 | 1.07 | 0.782 | 0.099 | -0.15 | 1.71 |
| Use of active comparator not available | -0.568 | 0.289 | -1.62 | 0.48 | -1.256 | 0.025 | -2.35 | -0.16 |
| Prevalence not available | 0.091 | 0.691 | -0.36 | 0.54 | -0.296 | 0.263 | -0.81 | 0.22 |
| Use of patient submission not available | 0.186 | 0.689 | -0.73 | 1.10 | 0.435 | 0.357 | -0.49 | 1.36 |
| Constant | -239.981 | . | . | . | -378.909 | 0.153 | -899.20 | 141.38 |

Note: Recommended technologies are the reference case

## Pooled analysis: sensitivity analysis of Pooled data set including 4 HTA bodies and ICER

**Table 8.4 Multivariate analysis of pooled sample of NICE, SMC, CVZ and HAS coverage decisions 2004-2009: base-case model results (n=977)**

| | Restricted vs. Recommended | | | | Not Recommended vs. Recommended | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.021 | 0.340 | -0.06 | 0.02 | -0.055 | 0.085 | -0.118 | 0.008 |
| RCT duration of follow-up | -0.002 | 0.364 | -0.0048 | 0.0017 | -0.003 | 0.085 | -0.0072 | 0.00046 |
| Use of active comparator in RCT | -0.522 | 0.015 | -0.94 | -0.10 | -0.895 | <0.001 | -1.353 | -0.436 |
| Clinical superiority demonstrated in RCT | -0.266 | 0.177 | -0.65 | 0.12 | -0.638 | 0.003 | -1.065 | -0.210 |
| ICER | -0.0000035 | 0.148 | -0.0000083 | 0.0000013 | 0.00000085 | 0.691 | -0.0000033 | 0.0000050 |
| Interaction term (no ICER within HAS decision-making | -0.000033 | 0.043 | -0.000065 | -0.000001 | -0.000036 | 0.038 | -0.000070 | -0.0000020 |
| Eligible patient population | 0.00000020 | 0.142 | -0.000000066 | 0.00000046 | 0.00000028 | 0.040 | 0.00000001 | 0.00000055 |
| Orphan designation status | -0.223 | 0.451 | -0.80 | 0.36 | -0.778 | 0.021 | -1.437 | -0.119 |
| Patient submission included | 0.436 | 0.159 | -0.17 | 1.04 | 0.769 | 0.016 | 0.141 | 1.396 |
| Number of technologies appraised simultaneously | 0.506 | 0.003 | 0.17 | 0.84 | -0.006 | 0.983 | -0.558 | 0.547 |
| National population size | 0.00000073 | 0.004 | 0.00000023 | 0.0000012 | 0.0000012 | <0.001 | 0.00000068 | 0.0000018 |
| Central nervous system | -0.086 | 0.789 | -0.72 | 0.55 | 0.196 | 0.562 | -0.468 | 0.860 |
| Eye | -1.630 | 0.025 | -3.06 | -0.20 | -1.279 | 0.073 | -2.679 | 0.121 |
| Malignancy/immunosuppresion therapy | -0.694 | 0.010 | -1.22 | -0.16 | -0.533 | 0.069 | -1.107 | 0.041 |
| Musculoskeletal and joint diseases | -0.280 | 0.397 | -0.93 | 0.37 | -0.900 | 0.027 | -1.697 | -0.103 |
| Obstetrics, gynaecology, and urinary-tract disorders | 2.329 | 0.032 | 0.20 | 4.46 | 1.681 | 0.150 | -0.606 | 3.968 |
| Respiratory system | 0.119 | 0.878 | -1.40 | 1.63 | 1.585 | 0.030 | 0.157 | 3.013 |
| Cardiovascular disease | -0.228 | 0.469 | -0.85 | 0.39 | -0.459 | 0.184 | -1.137 | 0.218 |
| Skin | -0.110 | 0.800 | -0.96 | 0.74 | -0.308 | 0.518 | -1.243 | 0.627 |
| NICE | -26.999 | 0.005 | -45.67 | -8.33 | -46.133 | <0.001 | -66.803 | -25.463 |
| SMC | 9.282 | 0.001 | 3.71 | 14.86 | 15.745 | <0.001 | 9.594 | 21.895 |
| HAS | -29.062 | 0.009 | -50.90 | -7.23 | -51.835 | <0.001 | -75.910 | -27.760 |
| Constant | -12.081 | 0.004 | -20.23 | -3.93 | -20.094 | <0.001 | -29.113 | -11.075 |

Note: Recommended technologies are the reference case. Multinomial logistic regression, pseudo R-squared: 0.26.

**Table 8.5 Base case 2: BASE CASE WITHOUT HAS, with ICER**

| | Restricted vs. Recommended | | | | Not Recommended vs. Recommended | | | |
|---|---|---|---|---|---|---|---|---|
| Variables | Log Odds | P value | 95% Conf. Interval | | Log Odds | P value | 95% Conf. Interval | |
| Number of Trials | -0.024 | 0.304 | -0.069 | 0.021 | -0.057 | 0.116 | -0.129 | 0.014 |
| RCT duration of follow-up | -0.002 | 0.498 | -0.006 | 0.003 | -0.004 | 0.112 | -0.010 | 0.001 |
| Use of active comparator in RCT | -0.684 | 0.007 | -1.184 | -0.184 | -1.382 | 0.000 | -1.949 | -0.815 |
| Clinical superiority demonstrated in RCT | -0.596 | 0.010 | -1.048 | -0.143 | -0.716 | 0.007 | -1.233 | -0.200 |
| ICER | -0.0000028 | 0.275 | -0.0000078 | 0.0000022 | 0.00000089 | 0.713 | -0.0000039 | 0.0000056 |
| Size of eligible patient population | -0.00000028 | 0.447 | -0.0000010 | 0.0000004 | -0.00000003 | 0.916 | -0.0000005 | 0.0000004 |
| Orphan designation | 0.075 | 0.846 | -0.686 | 0.837 | -0.038 | 0.927 | -0.853 | 0.777 |
| Patient submission | 0.410 | 0.188 | -0.201 | 1.020 | 0.639 | 0.049 | 0.003 | 1.274 |
| Number of technologies appraised simultaneously | 0.550 | 0.003 | 0.185 | 0.915 | 0.295 | 0.259 | -0.217 | 0.806 |
| National population size | 0.000 | 0.090 | 0.000 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 |
| Central nervous system | -0.127 | 0.713 | -0.804 | 0.550 | 0.323 | 0.380 | -0.398 | 1.044 |
| Eye | -1.992 | 0.077 | -4.204 | 0.219 | -1.656 | 0.191 | -4.139 | 0.828 |
| Malignancy/immunosuppresion therapy | -0.541 | 0.085 | -1.157 | 0.075 | 0.012 | 0.973 | -0.665 | 0.689 |
| musculoskeletal and joint diseases | 0.510 | 0.223 | -0.311 | 1.330 | -0.059 | 0.912 | -1.116 | 0.997 |
| obstetrics, gynaecology, and urinary-tract disorders | 2.515 | 0.022 | 0.370 | 4.659 | 1.761 | 0.150 | -0.634 | 4.156 |
| Respiratory system | 0.082 | 0.921 | -1.535 | 1.699 | 1.896 | 0.012 | 0.417 | 3.376 |
| Cardiovascular disease | 0.227 | 0.560 | -0.536 | 0.990 | 0.118 | 0.793 | -0.766 | 1.002 |
| Skin | -0.272 | 0.580 | -1.236 | 0.692 | -0.477 | 0.423 | -1.643 | 0.689 |
| NICE | -32.906 | 0.095 | -71.548 | 5.736 | -82.219 | 0.004 | -138.557 | -25.880 |
| SMC | 11.116 | 0.058 | -0.376 | 22.609 | 26.578 | 0.002 | 9.883 | 43.273 |
| Constant | -14.623 | 0.090 | -31.507 | 2.261 | -36.103 | 0.004 | -60.565 | -11.641 |